

(19) World Intellectual Property Organization
International Bureau



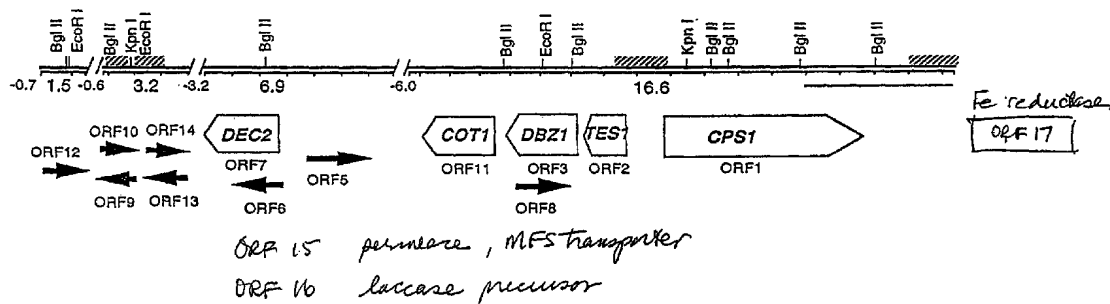
(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42444 A2

- (51) International Patent Classification⁷: **C12N 15/00**
- (21) International Application Number: PCT/US01/43381
- (22) International Filing Date:
21 November 2001 (21.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/252,649 22 November 2000 (22.11.2000) US
60/252,732 22 November 2000 (22.11.2000) US
- (71) Applicants (for all designated States except US): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 (CH). **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; 20 Farnwood Drive, Suite 105, Ithaca, NY 14850 (US).
- (71) Applicants and
(72) Inventors: **YODER, Olen** [US/US]; 4939 Concannon Court, San Diego, CA 91230 (US). **TURGEON, Barbara, G.** [CA/US]; 4939 Concannon Court, San Diego, CA 92130 (US). **LU, Shen-wen** [US/US]; 604 Winston Court, Apt. 4, Ithaca, NY 14850 (US).
- (74) Agent: **VIKSINIS, Ann, S.**; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: FUNGAL GENE CLUSTER ASSOCIATED WITH PATHOGENESIS



(57) Abstract: Methods to identify orthologs of fungal CPS1 genes as well as fungal iron reductase and permease/and or MFS transporter genes, and uses thereof are provided.

FUNGAL GENE CLUSTER ASSOCIATED WITH PATHOGENESIS

Cross-Reference to Related Applications

5 This application claims the benefit of the filing date of U.S. application Serial No. 60/252,649, filed on November 22, 2000, and U.S. application Serial No. 60/252,732, filed November 22, 2000, under 35 U.S.C. § 119(e), the disclosures of which are incorporated by reference herein.

Statement of Government Rights

10

The present invention was made with support from the United States Government (grant No. 96-35303-3198 from the USDA/NRI). The United States Government may have certain rights in the invention.

Field of the Invention

15

The present invention relates to DNA molecules comprising fungal, e.g., *Cochliobolus heterostrophus*, genes from a peptide synthetase gene cluster, e.g., an iron reductase and/or a permease or major facilitator superfamily transporter, and uses thereof.

20

Background of the Invention

There are approximately 30 species included in the genus *Cochliobolus*, nearly all of which are pathogens of wild grasses or cereals (Yoder et al., In: The Mycota Vol. 5: Plant Relationships, Part A, Berlin: Springer-Verlag, Carroll, eds., pp. 145-166 (1997)). *Cochliobolus heterostrophus* represents the most
25 widely distributed species in the genus and can be found in many tropical and subtropical areas in the world. As a natural pathogen of corn, *C. heterostrophus* causes a disease frequently called leaf spot of maize in the old literature (Drechsler, J. Agr. Res., 31:701 (1925); Drechsler, Phytopathol., 24:953 (1934);
30 Yu, "Studies on *Helminthosporium maydis*," 36:327 (1952)). In the United States, *C. heterostrophus* is usually found in the warmer southern states, thus, the

disease is commonly known as Southern Corn Leaf Blight (Hooker, Ann. Rev. Phytopathol., 12:167 (1974)). For many years, Southern Corn Leaf Blight was only known as an endemic disease and was not considered to be major economic importance in the United States. But in 1970, it suddenly broke into a severe epidemic that destroyed 15% of the U.S. corn crop and caused losses estimated at more than \$1 billion. This serious damage made Southern Corn Leaf Blight one of the most widely known crop diseases in the U.S.

Prior to the outbreak of the disease, only one race of *C. heterostrophus* (race O) was known in the field. In late 1969 when the disease became an epidemic, a new race of the fungus was identified from infected corn leaves collected in severely diseased areas. It was soon designated as race T because of its high virulence on T-cytoplasm corn and the ability to produce a phytotoxin called T-toxin, which specifically affects T-corn. In contrast, race O does not produce T-toxin and is mildly virulent on both T-cytoplasm and N-cytoplasm (normal cytoplasm) corn (Hooker et al., Plant Dis. Repr., 54:1109 (1970); Scheifele, "Cytoplasmically Inherited Susceptibility to Diseases Related to Cytoplasmically Controlled Pollen Sterility in Maize," 25:110 (1970); Smith et al., Plant Dis. Rep., 54:819 (1970); Yoder et al., Phytopathology, 65:273 (1975); Yoder, In: Biochemistry and Cytology of Plant Parasite Interaction, New York, New York:Elsevier, Tomiyama, eds., pp. 16-24 (1976); Yoder, Ann. Rev. Phytopathol., 18:103 (1980)). T-cytoplasm stands for Texas male sterile cytoplasm, a unique cytoplasm with a trait for maternally inherited male sterility, characterized by the failure to produce pollen (Levings, Science, 250:942 (1990)). T-cytoplasm corn was widely used for hybrid seed production and breeding to avoid hand or mechanical emasculation in the 1950s and the 1960s. It was the coexistence of large acreages of intensively planted T-cytoplasm corn and the sudden appearance of race T of *C. heterostrophus* that resulted in the epidemic of the disease in 1970. This discovery first opened the door to understanding pathogenesis by *C. heterostrophus*.

Early genetic analysis suggested that both T-toxin production and high virulence on T-cytoplasm corn are controlled by a single genetic locus defined as

Tox1 (Leach et al., Physiol. Plant Pathol., 21:327 (1982)). This was demonstrated by crosses between race T and race O in which only parental phenotypes segregated in a 1:1 ratio (Tox+:Tox-); all T-toxin producing progeny are highly virulent on T-cytoplasm corn while all T-toxin nonproducing progeny are weakly virulent (Yoder et al., 1975, *supra*; Leach et al., 1982, *supra*). Further investigation by comparison of electrophoretic karyotypes and chromosome-specific DNA hybridizations indicated that *Tox1* is tightly linked to a reciprocal translocation breakpoint and is associated with as much as a megabase of DNA (mostly highly repeated and A+T-rich) that is missing in race O (Bronson, Genome, 30:12 (1988); Tzeng et al., Genetics, 130:81 (1992); Chang et al., Genome, 39:549 (1996)). Surprisingly, recent analysis of several *Tox* mutants revealed that *Tox1* is not a single locus but rather two loci, each on a different translocated chromosome (Yoder et al., In Host-Specific Toxin: Biosynthesis, Receptor and Molecular Biology, Tottori, Japan: Faculty of Agriculture, Tottori Univ., Kohmoto, eds., pp. 23-32 (1994); Turgeon et al., Can. J. Bot., 73:S1071 (1995)). These two *Tox1* loci have been designated *Tox1A* and *Tox1B* (Yoder et al., 1997, *supra*). Two genes *PKS1* and *DECI* have been cloned from the two loci respectively; both are required for biosynthesis of T-toxin and are found only in race T isolates of *C. heterostrophus* (Yang, "The Molecular Genetics of T-Toxin Biosynthesis by *Cochliobolus heterostrophus*," Ph.D. Thesis, Cornell University (1995); Yang et al., Plant Cell, 8:2139 (1996); Rose et al., 8th Int. Symp. Mol. Plant-Microbe Int., Knoxville, p. J-49 (1996)).

Genetic analysis also suggested that T-toxin is required by *C. heterostrophus* for its high virulence on T-cytoplasm corn. This hypothesis was first tested by the generation of induced T-toxin deficient mutants using different mutagenesis procedures. All mutants with a tight *Tox*⁻ phenotype cause disease symptoms that are indistinguishable from those caused by race O when tested on both T and N-cytoplasm corn, suggesting that T-toxin is indeed a virulence factor (Yang et al., 1992; Lu et al., Proc. Natl. Acad. Sci. USA, 91:12649 (1994); Rose et al. (1996), *supra*). This conclusion was firmly supported by the site-specific disruption of the *PKS1* or *DECI* in the wild type race T genome;

disruptants lost the ability to produce T-toxins and caused race O type symptoms on both T-corn and N-corn (Yang et al., 1996, *supra*; Rose et al., 1996, *supra*). These experiments have given a very clear resolution for the role of T-toxin in pathogenesis. They also implied that pathogenesis by *C. heterostrophus* must involve additional pathogenicity factors because race O which does not produce T-toxin and race T-derived *Tox⁻* mutants are effective pathogens on corn.

A number of fungal molecules have been identified as general pathogenicity or virulence factors in several plant-pathogenic fungi (Yoder et al., *J. Genet.*, 75:425 (1996)). These include potential penetration factors such as melanin (Guillen et al., *Fungal Genet. Newsl.*, 41:41 (1994)), cutinase (Oeser et al., *Mol. Plant-Microbe Int.*, 7:282 (1994)) and polygalacturonase and xylanase (Lyngholm et al., *Fungal Genet. Newsl.*, 42:46 (1995)) or possible mechanisms involved in colonization such as phytotoxin detoxification (Schafer et al., *Science*, 246:247 (1989)) or components of signal transduction pathways. Although *C. heterostrophus* is known to produce a nonhost specific toxin called ophiobolin (or cochliobolin), a C₂₅ sesterterpenoid compound, which is toxic to many organisms, including plants, bacteria, fungi and nematodes, there is no evidence that ophiobolins are involved in pathogenesis by *C. heterostrophus* or other phytopathogenic fungi. No other pathogenesis-related toxins have been isolated from *C. heterostrophus* so far, but studies on closely related *Cochliobolus* species and other phytopathogenic fungi suggest that pathogenesis by this group of fungi also involves peptide toxins.

Four peptide phytotoxins (victorin, HC-toxin, AM-toxin, and enniatins) have been characterized as pathogenicity or virulence factors. They are all small cyclic peptides (4-6 residues), containing unusual amino acids or hydroxy acids, and they can be either host specific or non-host specific in terms of plant toxicity. A number of peptide phytotoxins are believed to be synthesized nonribosomally. Early in the 1960s, several biochemists working on the bacterial peptide antibiotics gramicidin and tyrocidine found that these polypeptides can be synthesized in RNAase-treated particle-free extracts of *Bacillus brevis* that are known to produce the same antibiotics; adding protein-

synthesis inhibitors to the extracts does not affect this process. This indicated the existence of a peptide biosynthetic system in which ribosomes and mRNAs are not needed. Further studies revealed that in this system, peptides are synthesized on a protein-template and this template itself is a multifunctional enzyme or a complex of several such enzymes, collectively called peptide synthetases, catalyzing the biosynthetic process (Laland et al., Essays in Biochemistry, 7:31 (1973); Lipmann, Adv. Microbiol. Physiol., 21:277 (1980)).

Peptide synthetases can catalyze biosynthesis of a variety of peptides. In terms of bioactivity, they can be antibiotics, enzyme inhibitors, plant or animal toxins and immunosuppressants (Stachelhaus et al., Journal of Biological Chemistry, 270:6163 (1995)). In terms of chemical structure, they can be either linear (i.e., ACV, the penicillin precursor and gramicidin) or cyclic (most are). The latter can be further classified into three subgroups: 1) The “standard” cyclic peptides (i.e., gramicidin S, tyrocidine, HC-toxin and cyclosporin); 2) cyclic lactones (i.e., destruxin); and 3) cyclic depsipeptides (i.e., beauvericin and enniatin). There have been over 300 different carboxy compounds that can be activated by peptide synthetases.

Although the first peptide synthetase, Gramicidin S synthetase, was purified and used for the cell-free synthesis of the peptide early in the 1960s (Tomino et al., Biochem, 6:2552 (1967)), the first bacterial peptide synthetase gene, *tycA*, which encodes the tyrocidine synthetase 1 in *B. brevis*, was not cloned until almost twenty years later (Marahiel et al., Mol. Gen. Genet., 201:1986 (1985)). Since then, more than twenty peptide synthetase genes have been reported for both bacteria and filamentous fungi, but only fourteen have complete nucleotide sequences published. All are larger than 3.3 kb and range between 3.3-19.5 kb for bacterial genes and 9.4-45.8 kb for fungal ones. Interestingly, all fungal peptide synthetase genes reported lack introns, even the cyclosporin A synthetase gene *simA*, which has a 45.8 kb of open reading frame (the largest genomic ORF so far recorded). Although biosynthesis of bacterial peptides differs from that of fungal ones in terms of the number of

multifunctional enzymes involved, the genes encoding these enzymes are similar to each other in both function and structure.

Comparison of nucleotide sequences reveals one or more highly conserved regions at certain positions in each peptide synthetase gene. These regions formerly called “amino acid activating domains” (Stachelhaus et al., 1995, *supra*), now called “amino acid activating modules” (Marahiel, Chem. Biol., 4:561 (1997)) consist of a set of domains (formerly called “modules”) believed to have specific functions such as recognition, activation and thioesterification of individual constituent amino or hydroxy acids, and in some cases methylation and racemation for modification of certain residues before incorporation into the peptide chain (Stachelhaus et al., 1995, *supra*). The most convincing evidence supporting this assignment is that in most cases, the number of conserved functional units in each gene or gene cluster is equal to the number of amino acids in the respective peptide. This one-for-one match is very clear between three of four fungal peptides and their biosynthetic genes. The total number of modules in three of four bacterial gene clusters also matches the number of amino acids in the respective peptides.

Sequence alignment of amino acid-activating modules reveals strictly conserved sequence motifs that contain active residues for module functions. These motifs are called “core sequences” (Marahiel, FEBS Lett., 307:40 (1992)). A minimal amino acid-activating module must contain six core sequences, whose functions (except for core 1) have been proposed based on mutational analysis of several peptide synthetases. Core sequences 1-5 are grouped into an amino acid adenylation domain and core 6 is a thioester formation domain (Figure 1A). All bacterial peptide synthetase genes contain “type I modules,” the minimal amino acid activating modules which were previously called “type I domains” (Stachelhaus et al., 1995, *supra*). Two fungal genes, *acvA* and *HTS1* also have this modular structure. In addition to the type I module, two fungal genes, *esyn1* and *simA*, contain type II modules, in which an insertion (about 400 amino acids) is found between cores 5 and 6 of a normal type I module. This region contains a motif (VLE/DXGXGXG; SEQ ID NO:1), highly conserved in

S-adenosyl-methionine (SAM)-dependent methyltransferases, hence, it is referred to as a N-methylation domain (Figure 1A). Additional evidence for methyltransferase activity of this module is that the number and position of type II modules in *esyn1*, and *simA* exactly match that of N-methylated amino acids in ennatin and cyclosporin sequences (Figure 1B).

Although the modular structure described above is highly conserved among most peptide synthetase genes, some variations have been found in the latest cloned peptide synthetase gene *safB*, which is the first gene in the saframycin Mx1 synthetase gene cluster (Pospiech et al., Microbiology, 141:1793 (1995)). *safB* contains two type I amino acid activating modules. One module has all six highly conserved core sequences, but another, believed to activate alanine (the first amino acid in the linear tetrapeptide precursor of saframycin Mx1), lacks core 5 and has a weakly conserved core 1 (Pospiech et al., Microbiology, 142:741 (1996)) (Figure 1A). This suggests that some of the motifs in the amino acid adenylation domain are dispensable or not critical for domain function. It also raises the possibility that other variations might be found in yet unknown peptide synthetase genes.

Although *C. heterostrophus* has been a model eukaryotic plant pathogen since the 1970s, most molecular genetic analyses conducted in this system have focused on production of the polyketide T-toxin by race T isolates of the fungus. Solid evidence now indicates that T-toxin is a host-specific virulence factor in Southern Core Leaf Blight (Yoder et al., J. Genet., 75:425 (1996); Yoder et al., 1997). It is clear, however, that *C. heterostrophus* needs additional factors, presumably general factors for pathogenesis to corn plants, since race O, which does not produce T-toxin, can be an effective corn pathogen. Attempts to identify additional general factors required by *C. heterostrophus* for pathogenesis have been unsuccessful.

Thus, what is needed is the isolation and characterization of additional fungal genes that control the biosynthesis of novel fungal molecules associated with pathogenesis, i.e., genes which are potential targets for the design of

products that might interfere with the infection process, and vertebrate fungal orthologs of fungal peptide synthetase genes.

Summary of the Invention

5 The invention generally relates to an isolated nucleic acid molecule (polynucleotide), e.g., DNA or RNA, comprising a nucleic acid segment which encodes a gene product related to pathogenesis. In one embodiment of the invention, fungal genes which are related to pathogenesis are identified. An advantage of the present invention is that the genes described herein provide the

10 basis to identify a novel fungicidal or mycocidal mode of action which permits rapid discovery of novel inhibitors of gene products that are useful as fungicides or mycocides. In addition, the invention provides isolated genes or gene products from fungi for assay development for inhibitory compounds with fungicidal or mycocidal activity, as agents which inhibit the function or reduce or

15 suppress the activity of those gene products in fungi are likely to have detrimental effects on fungi, and are good fungicide or mycocide candidates. The present invention therefore also provides methods of using a polypeptide encoded by one or more of the genes of the invention or a cell expressing such a

20 polypeptide to identify inhibitors of the polypeptide, which can then be used as fungicides to suppress the growth of pathogenic fungi. Pathogenic fungi are defined as those capable of colonizing a host and causing disease. Examples of fungal pathogens include plant pathogens such as *Septoria trici*, *Ashbya gossypii*, *Stagenospora nodorum*, *Botryus cinera*, *Fusarium graminearum*, *Magnaporthe grisea*, *Cochliobolus heterostrophus*, *Colleototrichum*, *Ustilago maydis*,

25 *Erisyphe graminis*, plant pathogenic oomycetes such as *Pythium ultimum* and *Phytophthora infestans*, as well as dimorphic fungal pathogens including *Blastomyces*, e.g., *B. dermatitidis*, *Coccidioides*, *Histoplasma*, e.g., *H. capsulatum*, or *Paracoccidioides*, e.g., *P. brasiliensis*, *Loboa*, *Malassezia*, *Rhodotorrula*, *Blastoschizomyces*, *Trichosporon*, *Saccharomyces*, *Cryptococcus*

30 including *Cryptococcus neoformans*, as well as human pathogens such as *Candida albicans*, and other pathogenic *Candida*, e.g., *C. tropicalis*, *C.*

parapsolosis and *C. guettermondii*, *Coccidioides imitis*, and *Aspergillus fumigatus*, *Sporothrix schenckii*, pathogenic members of the Genera *Epidermophyton*, *Microsporum* and *Trichophyton*, *Cladosporium* (*Xylohypha*) *trichoides*, *Cladosporium bantianum*, *Penicillium marneffii*, *Exophiala* (5 *Wangiella*) *dermatitidis*, *Fonsecaea pedrosoi* and *Dactylaria gallopava* (*Ochroconis gallopavum*), and including mycogens. Preferred fungi for use with the agent identified by the method of the invention are *Ascomycota*.

In one embodiment of the invention, the invention relates to an isolated polynucleotide comprising a nucleic acid segment encoding an ortholog of a
 10 plant fungal CPS1, e.g., SEQ ID NO:3 from *Cochliobolus* which is a CoA ligase, or a nucleic acid segment encoding a gene product that modulates fungal iron metabolism, uptake, absorption of inorganic or organic ferric salts, e.g., a fungal iron reductase, permease or MFS transporter, e.g., a siderophore transporter, which genes maybe associated with *CPS1* in a gene cluster. As described herein
 15 below, a gene from *Coccidioides imitis* and *Candida* that is related to the *CPS1* gene of *Cochliobolus* was identified, e.g., a nucleic acid sequence comprising an open reading frame comprising SEQ ID NO:46 which encodes SEQ ID NO:47 or the complement thereof. The *CPS1* gene in *Cochliobolus* is present in a cluster of closely linked open reading frames, a cluster which is associated with
 20 virulence and/or pathogenicity, wherein CPS1 is representative of a novel class of adenylation domain-containing enzymes related to but distinct from nonribosomal protein synthetases (NRPSs). Thus, at least one of the genes in the cluster may control biosynthesis of a secondary metabolite (small molecule) that is required for or associated with fungal virulence and/or pathogenesis.
 25 Similarly, orthologs of the described *Cochliobolus* gene cluster, e.g., those in *Coccidioides* or *Candida*, may encode gene products that are required for or associated with fungal virulence. As also described hereinbelow, a *Cochliobolus* iron reductase (SEQ ID NO:49 encoded by SEQ ID NO:48) and a permease and/or MFS transport protein gene (SEQ ID NO:55 encoding SEQ ID NO:56)
 30 were identified that are closely linked to a *CPS1* peptide synthetase gene, e.g., a DNA molecule comprising SEQ ID NO:2 (GenBank accession no. AF332878)

encoding SEQ ID NO:3 (GenBank accession no. AAG53991), which is part of a gene cluster associated with virulence and/or pathogenicity.

Thus, at least one of the genes in the cluster may control biosynthesis of at least one secondary metabolite or other small molecule that is required for or associated with fungal growth, virulence and/or pathogenesis. The fungal produced siderophore may sequester iron from the environment or host to aid in fungal growth. *Pseudomonas aeruginosa* produces pigments that are likely associated with virulence, e.g., pyocyanin. A derivative of pyocyanin, pyochelin, is a siderophore that is produced under low iron conditions to sequester iron from the environment for growth of the pathogen. The competition for iron may have a deleterious effect on the host. Similarly, the *Cochliobolus* iron reductase or permease/transporter or other gene products associated with iron metabolism may compete with the host for Fe and so contribute to the pathogenicity of the fungus. Similarly, orthologs of the described genes in the *Cochliobolus* gene cluster in other fungi which infect plants or those that infects vertebrate animals may encode gene products that are required for or associated with fungal virulence including iron metabolism genes, e.g., genes associated with secretion of a toxin or siderophore.

Preferably, the nucleic acid segment is obtained or isolatable from a fungal gene which encodes a polypeptide which is substantially similar, and preferably has at least 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, amino acid sequence identity to, a polypeptide encoded by a nucleic acid sequence comprising any one of SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55, or a fragment (portion) thereof which encodes a partial length polypeptide having substantially the same activity of the full length polypeptide. Preferably, the activity of the partial length polypeptide is at least 50%, generally at least 60%, ordinarily at least 70%, preferably at least 80%, more preferably at least 90% and more preferably still at least 95% the activity as the full-length

polypeptide. Preferred partial length polypeptides have substantially the same activity as the corresponding full-length polypeptide.

Further provided is an isolated polynucleotide comprising a nucleic acid segment which is substantially similar, and preferably has 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, nucleotide sequence identity to, a nucleic acid sequence comprising an open reading frame comprising any one of SEQ ID NO: 46, SEQ ID NO:48, or SEQ ID NO:55.

Another aspect of the present invention, as described below, relates to a method for identifying inhibitors of the gene products encoded by the polynucleotides of the invention, which involves contacting the gene product or cell expressing the polynucleotide with agents that are potential inhibitor compounds, and selecting compounds which decrease the activity of the gene product and/or inhibit cell growth. In another embodiment, the invention relates to a method of imparting disease resistance to a plant or other organism by overexpression the CPS1 ortholog of the invention in the plant or other organism.

The nucleic acid molecules of the invention are preferably obtained or isolatable from a gene from fungi that infect vertebrates, including but not limited to mammals, e.g., livestock such as bovine, ovine, porcine, equine and avians such as turkey and chickens and domestic pets including avians, feline and canine, and humans, which genes are related to pathogenesis. For example, preferred nucleic acid molecules of the invention are obtained or isolatable from *Ascomycetes* (ascomycetes), and the agents of the invention are useful to treat infections due *Ascomycota* infection, based on the discovery of CPS1, its orthologs and related genes in the cluster, in various ascomycetes human (and plant) pathogens as disclosed herein. Within pathogenic *Ascomycetes*, the following groups are of interest: *Agyriales*, *Arthoniales*, *Ascosphaerales*, *Caliciales*, *Calosphaeriales*, *Capnodiales*, *Chaetothyriales* (black yeasts), *Cyttariales*, *Diaporthales*, *Dothideales*, *Elaphomycetales*, *Erysiphales* (powdery

- mildews), *Eurotiales* (green and blue mold), *Gyalectales*, *Halosphaeriales*, *Helotiales*, *Hypocreales*, *Laboulbeniales*, *Lecanorales*, *Lulworthiales*, *Melanommatales*, *Meliolales*, *Microascales*, *Myriangiales*, *Neolectales*, *Onygenales*, *Ophiostomatales*, *Ostropales*, *Patellariales*, *Pertusariales*,
5 *Pezizales*, *Phyllachorales*, *Pleosporales*, *Protomycetales*, *Pyrenulales*, *Rhytismatales*, *Saccharomycetes*, *Schizosaccharomycetales*, *Sordariales*, *Taphrinales*, *Teloschistales*, *Thelebolaceae*, *Umbilicariales*, *Xylariales*, anamorphic *Ascomycota*, unclassified *Ascomycota*, and *Ascomycota incertae sedis*.
- 10 Regarding *Ascomycetes* animal pathogens, preferred are pathogenic *Onygenales*, more particularly the anamorphic *Onygenales*, which includes *coccidioides*, and the *Onygenaceae* and its group *Ajellomyces*, which includes *Histoplasma* such as *Histoplasma capsulatum*, and *Blastomycoides* such as *Blastomycoides dermatitidis*. Also preferred are pathogenic *Saccharomycetes*,
15 more preferably *Saccharomycetales*, and even more preferably anamorphic *Saccharomycetales*, which includes *Candida* species. Also preferred are *Chaetothyriales*, more preferably *Herpotrichiellaceae*, even more preferably anamorphic *Herpotrichiellaceae*, and even more preferably *Exophiala*, which include the human-pathogenic organisms *Exophiala dermatitidis* and *Exophiala*
20 *jeanselmei*. Also preferred are the *Onygenales*, more preferably *Arthrodermataceae*, more preferably anamorphic *Arthrodermataceae*, and even more preferably *Trichophyton*, which contain *Trichophyton rubrum*. Another preferred group is Fungi incertae sedis, more preferably *Pneumocystidaceae*, and even more preferably *Pneumocystis*, which includes the human pathogen
25 *Pneumocystis carinii*. Yet another preferred group is *Eurotiales*, more preferred *Trichocomaceae*, even more preferred anamorphic *Trichocomaceae*, and yet even more preferred is *Aspergillus* species, which contains *Aspergillus avenaceus* and *Aspergillus fumigatus*. Another preferred group are those pathogenic fungi in *Pleosporales*, more preferably *Pleosporaceae*, yet more
30 preferably anamorphic *Pleosporaceae*, and even more preferably *Alternaria* species, which includes airborne *Alternaria alternata*. Also preferred is

Ascomycota incertae sedis, more preferably *Mycosphaerellaceae*, particularly the anamorphic *Mycosphaerellaceae*, and more preferably the species *Cladosporium*, which includes airborne human pathogens. Also preferred are anamorphic *Ascomycota*, more preferably the species *Helminthosporium*.

- 5 Within Onygenales are preferably anamorphic *Onygenales*, and more preferably the *Paracoccidioides* species, which includes *Paracoccidioides brasiliensis*. Also preferred are *Microascales*, more preferably *Microascaceae*, and even more preferably *Pseudallescheria* species, which includes *Pseudallescheria boydii*. Also preferred are *Ophiostomatales*, more preferably *Ophiostomataceae*, yet
- 10 more preferably anamorphic *Ophiostomataceae*, and more preferably *Sporothrix* species, including *Sporothrix schenckii*.

- The term “substantially similar”, when used herein with respect to a polypeptide means a polypeptide corresponding to a reference polypeptide, wherein the polypeptide has substantially the same structure and function as the
- 15 reference polypeptide, e.g., where the only changes in amino acid sequences are those which do not affect the polypeptide function. When used for a polypeptide or an amino acid sequence, the percentage of identity between the substantially similar and the reference polypeptide or amino acid sequence is at least 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%,
- 20 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, wherein the reference polypeptide comprises SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56. One indication that two polypeptides are substantially similar to each other is that an agent, e.g., an antibody, which specifically binds to one of the polypeptides,
- 25 specifically binds to the other.

- In its broadest sense, the term “substantially similar”, when used herein with respect to a nucleotide sequence or nucleic acid segment, means a nucleotide sequence or segment corresponding to a reference nucleotide sequence or nucleic acid segment, wherein the corresponding sequence encodes a
- 30 polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence or nucleic acid

segment. The term “substantially similar” is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence is at least 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, preferably wherein the reference sequence comprises SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55 or the complement thereof. Sequence comparisons may be carried out using a Smith-Waterman sequence alignment algorithm (see e.g., Waterman, Introduction to Computational Biology: Maps, sequences and genomes, Chapman & Hall, London (1995) or <http://www.htousc.edu/software/seqaln/index.html>). The local S program, version 1.16, is preferably used with following parameters: mat:1, mismatch penalty: 0.33, open-gap penalty:2, extended-gap penalty:2. Further, a nucleotide sequence that is “substantially similar” to a reference nucleotide sequence hybridizes to the reference nucleotide sequence under moderate, stringent, or very stringent, hybridization conditions, e.g., in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

Thus, the invention also includes recombinant nucleic acid molecules which have been modified so as to comprise codons other than those present in the unmodified sequence or have been modified by shuffling. The recombinant nucleic acid molecules of the invention include those in which the modified codons in the unmodified sequence, as well as those that specify different amino

acids, i.e., they encode a variant polypeptide having one or more amino acid substitutions relative to the polypeptide encoded by the unmodified sequence.

The invention further includes a nucleotide sequence which is complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with the nucleic acid molecules of the invention as well as RNA which is encoded by the nucleic acid molecules of the invention as well as RNA which is encoded by the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of the invention is preferably supported, e.g., on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of, e.g., between 55 and 70°C, in double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth strength SC containing 0.1% SDS.

Hence, the isolated nucleic acid molecules of the invention include orthologs of SEQ ID NO:46, SEQ ID NO:48 and SEQ ID NO:55, which includes orthologs of the polypeptides encoded therein. An ortholog is a gene from a different species that encodes a product having the same function as the product encoded by a gene from a reference organism. The encoded ortholog products likely have at least 68 to 70% (substantial) sequence identity to each other. Hence, one embodiment the invention includes an isolated polynucleotide comprising a nucleic acid segment encoding a polypeptide having at least 68 to 70% identity to a polypeptide encoded by SEQ ID NO:46, SEQ ID NO:48 or SEQ ID NO:55. Databases such as GenBank which can be accessed at <http://www.ncbi.nlm.nih.gov/>, may be employed to identify sequences related to those sequences. Alternatively, recombinant DNA techniques such as hybridization or PCR may be employed to identify sequences related to the sequences. Preferred orthologs include those from dimorphic fungal pathogens

including *Blastomyces*, e.g., *B. dermatitidis*, *Coccidioides*, *Histoplasma*, e.g., *H. capsulatum*, or *Paracoccidioides*, e.g., *P. brasiliensis*, *Loboa*, *Malassezia*, *Rhodotorrula*, *Blastoschizomyces*, *Trichosporon*, *Saccharomyces*, *Cryptococcus* including *Cryptococcus neoformans*, as well as human pathogens such as

5 *Candida albicans*, and other pathogenic *Candida*, e.g., *C. tropicalis*, *C. parapsilosis* and *C. guilliermondii*, *Coccidioides immitis*, and *Aspergillus fumigatus*, *Sporothrix schenckii*, pathogenic members of the Genera *Epidermophyton*, *Microsporum* and *Trichophyton*, *Cladosporium* (*Xylohypha*) *trichoides*, *Cladosporium bantianum*, *Penicillium marneffii*, *Exophiala*

10 (*Wangiella*) *dermatitidis*, *Fonsecaea pedrosoi* and *Dactylaria gallopava* (*Ochroconis gallopavum*), as well as other mycogens.

The invention also provides anti-sense nucleic acid molecules corresponding to the sequences described herein. Also provided are expression cassettes, e.g., recombinant vectors, and host cells, comprising the nucleic acid

15 molecule of the invention in which the nucleic acid segment is in either sense or antisense orientation. Also provided is a microarray, comprising one or more of the nucleic acid molecules of the invention or a portion thereof.

Owing to the dramatically increased incidence of life-threatening opportunistic fungal infections it is now clear that diseases of fungal infection

20 are of major importance. The rise in cases has been particularly apparent in transplant recipients and others who are immunocompromised, especially AIDS patients. Besides more serious infections associated with these vulnerable groups, superficial infections such as ringworm and thrush have also become more prevalent. Despite recognizing the importance of fungi as a cause of

25 disease in man and animals, many of the more serious fungal infections remain difficult to diagnose and treat. Thus, there is a continuing need to identify agents to treat fungal infections of vertebrates, including immunocompromised vertebrates, and complications thereof, e.g., pneumonia, flu-like illness, erythema nodosum, erythema marginatum, arthritis, multiple thin-walled chronic cavities,

30 miliary disease, bone and joint infection, skin disease, soft tissue abscesses, meningitis, oropharyngitis, oesophagitis, vaginitis, onychomycosis,

endophthalmitis, paronychia, and inflammation of the urinary tract, kidney, liver, brain, gastrointestinal tract, and lung.

Thus, another aspect of the present invention relates to a method for identifying inhibitors of the fungal vertebrate CPS1 ortholog, or fungal iron reductase or permease/MFS transporter of the invention. For example, genes
5 encoding products that are associated with virulence, and agents that bind to or otherwise alter or modulate the activity of that gene product, preferably agents that inactivate or decrease (reduce or inhibit) the activity of the gene product, can be identified. The method comprises contacting the gene product(s) or cells
10 which express the gene product(s) with an agent and then determining or detecting whether the agent binds to, or decreases the activity of, the gene product(s). Such an agent modulates or alters a phenotype of the gene product or cell, e.g., pathogenicity of a cell which expresses the gene product. Modulation or alteration encompasses an increase as well as a decrease in an activity,
15 preferably the modification or alteration in the activity of the gene product or cell having the gene product contacted with the agent is at least 10%, or at least 50%, relative to the activity in an untreated control. In particular, the methods are useful to identify agents that inhibit, reduce or suppress the activity of the polypeptide, e.g., by at least 10%, preferably at least 50%, relative to the activity
20 in an untreated control. Thus, the invention also provides agents identified by the methods of the invention. Preferred agents bind to, more preferably inhibit, the activity of a polypeptide of the invention, e.g., one encoded by a dimorphic fungal pathogen such as one from *Blastomyces*, *Coccidioides*, *Histoplasma* or *Paracoccidioides*, and includes pathogenic *Candida*, e.g., *C. albicans*, *C. tropicalis*, *C. parapsolosis* and *C. guettermondii*. The methods may employ
25 screening agents on wild type fungi and/or recombinant fungi, e.g., fungi which overexpress the polypeptide of interest or do not express that polypeptide, e.g., as a result of expression of antisense sequences or a gene knock out. If the agent is one encoded by DNA, the expression of that DNA in an organism susceptible to
30 the pathogen, e.g., a plant, may provide tolerance or resistance to the organism to the pathogen, preferably by inhibiting or preventing pathogen infection.

Methods of the invention may include stably transforming a susceptible organism of cell with one or more sequences which confer tolerance or resistance operably linked to a promoter capable of driving expression of that nucleotide in the cells of the organism.

5 Other uses for the nucleic acid molecules or polypeptides of the invention, include the use of the polypeptide to raise either polyclonal antibodies or monoclonal antibodies, e.g., antibodies specific for the polypeptide, to detect antibodies in the serum of a vertebrate, or primers or probes specific for the nucleic acid molecules, which can be employed in diagnostic assays for the
10 presence of the pathogen or for therapeutic purposes, and host cells comprising the nucleic acid molecules, e.g., in antisense orientation, or having a deletion in at least a portion of at least one the genes corresponding to the nucleic acid molecules of the invention. Also, given that the gene may encode a peptide synthetase (Watanabe et al., Chem. Biol., 3, 463 (1996)) the gene product may be
15 useful in therapy, e.g., as an anti-cancer agent, an antibiotic, or as an immunosuppressant.

 The agents identified by the methods of the invention may also be subjected to further assays to determine whether the agent is substantially non-toxic to a plant or vertebrate organism to be treated as well as the dose to be
20 administered to the vertebrate organism. For example, for *Coccidioides*, a murine model may be employed (see, Kirland et al., Infect. Immun., 40: 912 (1983)). This model may also be used for screening for an agent of the invention. Further, the agents identified by the methods of the invention, e.g., those which are non-toxic to a plant or vertebrate to be treated, are useful in
25 methods of preventing or treating a disease or disorder associated with fungal infection, including superficial, subcutaneous or systemic infections. The method comprises administering to a vertebrate or plant in need of such treatment, e.g., a vertebrate that is immunocompromised, an amount of an agent of the invention effective to inhibit or prevent fungal or mycogen infection or
30 growth. For example, humans and non-human animals including livestock and domestic pets may be treated with the agents of the invention, e.g., livestock

such as bovine, ovine, porcine, equine and avians such as turkey and chicken and domestic pets including avians, felines and canines. Preferably, the agents are administered topically to a mammal such as a human. Preferred plants include cereals, for example, corn, alfalfa, sunflower, rice, Brassica, canola, soybean, 5 barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat millet, and tobacco.

Moreover, the agents of the invention may be used in conjunction with other therapeutic agents, e.g., fungicides, mycosides, and vaccines, including amphotericin B and azoles. In addition, the agents may be employed to treat 10 sources of fungal contamination, such as the soil or surface areas or materials on which fungi can survive and/or proliferate. Thus, the agents may be contacted with soil or other surfaces that come in contact with vertebrates. Although this contacting may not eliminate the fungus, it may reduce the risk of airborne dissemination of the fungus or its spores.

15 Also provided is a computer readable medium having stored thereon a nucleic acid sequence that is substantially similar to any one of SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55 or the complement thereof, and a computer system comprising a processor and data storage device wherein said data storage device has stored thereon a nucleic acid sequence that is substantially similar to 20 any one of SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55 or the complement thereof. Preferably, the computer system comprises an identifier which identifies features in said sequence. Further provided is a database comprising at least one nucleotide sequence in computer readable form wherein said nucleotide sequence is substantially similar to any one of SEQ ID NO:46, SEQ ID NO:48, 25 SEQ ID NO:55, or the complement thereof. The database, for example, carries out functions comprising determining homology, aligning sequences, adjusting sequence alignments, assembling sequences having overlapping sequence, predicting gene sequence, predicting intron borders, identifying motifs, identifying domains, identifying untranslated regulatory sequences, identifying 30 putative sequencing errors, carries out functional genomics analyses, or carries out shuffling of nucleotide sequences.

The invention also provides a method for generating nucleotide sequences encoding polypeptides having at least one region of homology to SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55, or the complement thereof. The method comprises shuffling an unmodified nucleotide sequence which is

5 identical or substantially identical to SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55, or the complement thereof. The resulting shuffled nucleotide sequence is expressed and a gene product encoded thereby is selected for altered activity as compared to the activity in a polypeptide encoded by SEQ ID NO:46, SEQ ID NO:48, or SEQ ID NO:55. A DNA molecule comprising a shuffled nucleotide

10 sequence obtainable or produced by the method is also provided. In one embodiment, the shuffled DNA molecule encodes a polypeptide having enhanced tolerance to an inhibitor of the polypeptide encoded by SEQ ID NO:46, SEQ ID NO:48, or SEQ ID NO:55. The shuffled DNA molecule may be operably linked to a promoter to form a chimeric molecule which is introduced

15 to a host cell, e.g., a plant cell.

Brief Description of the Figures

Figure 1 provides the structure of amino-acid activating modules identified in peptide synthetase genes (adapted from Stachelhaus and Marahiel, J. Biol. Chem., 270, 6163, 1995; Stachelhaus and Marahiel, FEMS Microbiol. Lett., 125, 3, 1995; Pospiech 1995, *supra*; Marahiel, 1997, *supra*). Figure 1A shows the domain arrangements in two types of modules. Structural variations in the first module (safB1) of the gene *safB* are also indicated below type I. Figure 1B shows the correlation between module types and the nature of residues

20 in two fungal peptides. Open box: type I module; filled box: type II module. Each peptide sequence is given below.

Figure 2 is a restriction map of the cloned sequences surrounding the tagged site. A 11.3 kb genomic region (thick line) was cloned and completely sequenced. The original REMI insertion point in the mutant R.C4.2696 is

30 indicated by a vertical arrow. The asterisks indicate two targeted integration sites in the wild type genome. Two open reading frames (in opposite directions),

ORF1 (*CPS1*, 5.4 kb) and ORF2 (*TES1*, 1.1 kb) are indicated by open boxes below the map (the positions of putative introns are indicated by vertical bars). Locations of seven overlapping plasmid clones used for sequencing are indicated by thin lines on the top of the map (filled triangles represent the vector sequence in each clone). Sequencing strategy is indicated by arrow above each clone line.

Figures 3A-C are schematic representations which show the characterization of modular structure of *CPS1*. Peptide synthetase and thioesterase are indicated by open boxes; shaded boxes inside indicate functional domains and modules; vertical bars in the shaded boxes indicate highly conserved core sequences. Figure 3A illustrates the general structure of bacterial and fungal peptide synthetases (adapted from Marahiel, 1997, *supra*). A peptide synthetase gene cluster is shown on the top. There can be one or more amino acid activating module (cyclosporine synthetase has 11) in each protein; some peptide synthetases have thioesterase domains (TE), which can be either integrated into modules or encoded by a separate gene. Each synthetase can have type I, type II or both modules. A type I (minimal) module is enlarged to show organization of core sequences and domains. Some peptide synthetases also have condensation or epimerization domains. Figure 3B illustrates the organization of saframycin Mxl synthetase containing 4 amino acid activating modules (Pospiech et al., 1996, *supra*). SafB1 from the first module is enlarged. Core sequences 1 and 5 in safB1 are weakly conserved (indicated by dashed vertical bars). The remaining domains are typical of type I as shown in Figure 3A. SafC is a putative O-methyltransferase. Figure 3C illustrates the organization of *CPS1*. Sequence analysis revealed two amino acid activating modules (CPS1A and CPS1B), both of which have high similarity to safB1 except that core 2 is weakly conserved. A thioesterase domain is found at the C-terminal region of CPS1B. Three vertical arrows indicate the positions of targeted gene disruptions in the wild type genome that yielded the mutant phenotype. TES1 is a thioesterase encoded by a separate gene (TES1).

Figures 4A-C depict DNA gel blots showing DNA-DNA hybridization of ChCPS1 to other fungal genera and species. (A) *Cochliobolus* species (1-17); C.

heterostrophus race T, race O; *C. carbonum* race 1, race 2; *C. victoriae* isolates FI3, HvW; *C. bicolor*, *C. dactyloctenii*, *C. chloridis*, *C. homomorphus*, *C. intermedius*, *C. melinidis*, *C. melinidis*, *C. peregrinensis*, *C. perotidis*, *C. ravenelii* and *C. sativus*. (B) Other Ascomycete genera (1-14): *C. carbonum* race1 (control), *Setosphaeria rostrata*, *Stemphylium* spp., *Pyrenophora tritici* *repentis*, *Bipolaris sacchari*, *Alternaria* spp., *A. solani*, *Nectria haematococca*, *Fusarium oxysporum*, *Glomerella* spp. *Magnaporthe grisea*, *F. moniliforme*, *F. moniliforme* (repeat) and *A. solani* (repeat). (C) *Candida albicans* compared to *C. heterostrophus* and closely related species (1-7): *C. heterostrophus* race T, *Bipolaris sacchari*, *Setosphaeria rostrata*, *Stemphylium* spp., *Pyrenophora tritici* *repentis*, *Alternaria* spp. and *Candida albicans* (arrowhead). Genomic DNAs were digested with *Hind*III (A, lanes 1-17; B, lanes 1-11; C, lanes 1-7), *Xho*I (B, lanes 12 and 14) or *Bgl*II (B, lane 13) and probed with the 3.2 kb fragment of CPS1 at high stringency. Weak signals in lanes 3 and 17 (panel A) are due to insufficient DNA loading (confirmed by a repeat experiment).

Figures 5A-B show similarity of the cloned CPS1 homologs to *C. heterostrophus* CPS1. (A) Structural comparison of the four CPS1 homologs to ChCPS1 (As = *Alternaria solani*; Pt = *Pyrenophora teres*; Fg = *Fusarium graminearum*; Ci = *Coccidioides immitis*). ORFs are indicated by the open boxes; shaded boxes inside indicate functional domains; vertical bars indicate conserved motif sequences found in nonribosomal peptide synthetases (NRPS) as defined by Stachelhaus and Marahiel (Stachelhaus and Marahiel, 1995, *supra*; Marahiel, 1997, *supra*) (dashed bars indicate weak conservation). The black bulbs indicate the position of putative introns. Cores 1-5: adenylation; core 6: thiolation; TE: thioesterase. The distance between core sequences is not drawn in exact scale. The name of proteins is on the left of ORF box and the number of amino acids on the right. The unidentified regions of AsCPS1, PtCPS1 and CiCPS1 are indicated by dash-lined boxes. The similarity to ChCPS1 (in the overlapping region only) is given in the parentheses under the protein names in the order: nucleotide identity/ amino acid identity/ amino acid similarity. The positions of the ChCPS1 amino acids 220 and 1040 (corresponding to the first

and the last amino acid of CiCPS1) are indicated by open arrows; the positions 511 and 1269 (to the first and the last amino acids of AsCPS1 and PtCPS1) are indicated by filled triangles. (B) Amino acid alignment of the four CPS1 homologs to ChCPS1. 530 amino acids aligned to the amino acids 511-1040 of ChCPS1 (SEQ ID NO:186) are shown (SEQ ID NOs: 51-54). The identical residues are in uppercase and the similar residues in lowercase. Consensus of sequences similar to the typical NRPS signature motifs is underlined. The putative cyclization domain motif “DXXXXD/ EXXS/ A” (SEQ ID NO:60) is underlined.

Figure 6 shows the results of a BLAST search using FgCPS1 (SEQ ID NO:41) as the query sequence.

Figure 7A shows the results of a BLAST search using CiCPS1 (SEQ ID NO:47) as the query sequence.

Figure 7B shows an alignment of amino acid sequence of FgCPS1 (SEQ ID NO:41), AsCPS1 (SEQ ID NO:43), PtCPS1 (SEQ ID NO:45), CiCPS1 (SEQ ID NO:47), and ChCPS1 (SEQ ID NO:3).

Figures 8A-C show the sequencing strategy (A), restriction map (B), genome organization (C) for the *ChCPS1* gene cluster. SEQ ID NO:59 represents the sequence of genes clustered near *ChCPS1*. SEQ ID NO:187 and 188 represent the DNA corresponding to and amino acid sequence encoded by ORF 16, respectively. SEQ ID NO:189 and 190 represent the DNA corresponding to and amino acid sequence corresponding to ORF 10, respectively. SEQ ID NO:191 and 192 represent the DNA corresponding to and amino acid sequence encoded by ORF 11, respectively. SEQ ID NO:193 and 194 represent the DNA corresponding to and amino acid sequence encoded by ORF 12, respectively. SEQ ID NO:195 and 196 represent the DNA corresponding to and amino acid sequence encoded by ORF 13, respectively. SEQ ID NO:197 and 198 represent the DNA corresponding to and amino acid sequence encoded by ORF 14, respectively. SEQ ID NO:199 and 200 represent the DNA corresponding to and amino acid sequence encoded by ORF 3, respectively. SEQ ID NO:201 and 202 represent the DNA corresponding to and

amino acid sequence encoded by ORF 5, respectively. SEQ ID NO:203 and 204 represent the DNA corresponding to and amino acid sequence encoded by ORF 6, respectively. SEQ ID NO:205 and 206 represent the DNA corresponding to and amino acid sequence encoded by ORF 7, respectively. SEQ ID NO:207 and 208 represent the DNA corresponding to and amino acid sequence encoded by ORF 8, respectively. SEQ ID NO:209 and 210 represent the DNA corresponding to and amino acid sequence encoded by ORF 9, respectively.

Figure 9A shows the results of a BLAST search using SEQ ID NO:49 (an iron reductase encoded by SEQ ID NO:48) as the query sequence.

Figure 9B shows an alignment of amino acid sequence of a *Cochliobolus* iron reductase (SEQ ID NO:49) and a *S. cerevisiae* reductase (SEQ ID NO:184).

Figure 9C illustrates a DNA comprising SEQ ID NO:48 (SEQ ID NO:211).

Figure 9D illustrates the amino acid sequence (SEQ ID NO:212) encoded by SEQ ID NO:211.

Figure 10 shows the results of a BLAST search using the polypeptide (SEQ ID NO:56) encoded by SEQ ID NO:55 (a *Cochliobolus* permease and/or MFS transporter) as the query sequence.

Detailed Description of the Invention

Definitions

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated.

Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzner et al., Nucl. Acids Res., 19:508 (1991); Ohtsuka et al., JBC, 260:2605 (1985); Rossolini et al., Mol. Cell. Probes, 8:91 (1994). Although nucleotides are usually joined by phosphodiester linkages, polymeric nucleotides joined by peptide linkages (peptide nucleic acids) are also included (Nielsen and Egholm, Peptide Nucleotide Acids: Protocols and Applications, Horizon Scientific Press, Wymondham, Norfolk UK, 1999). A “nucleic acid fragment” is a fraction of a given nucleic acid molecule. Deoxyribonucleic acid (DNA) in the majority of organisms is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. The term “nucleotide sequence” refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The terms “nucleic acid”, “nucleic acid molecule”, “nucleic acid fragment” or “nucleic acid sequence or segment” may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an “isolated” or “purified” DNA molecule or an “isolated” or “purified” polypeptide is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an “isolated” or “purified” nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an “isolated” nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the

genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein-of-interest chemicals. Fragments and variants of the disclosed nucleotide sequences and proteins or partial-length proteins encoded thereby are also encompassed by the present invention.

By "fragment" or "portion" is meant a full length or less than full length of the nucleic acid sequence encoding, or the amino acid sequence of, a polypeptide or protein. Alternatively, fragments or portions of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments or portions of a nucleotide sequence may range from at least about 6 nucleotides, about 9, about 12 nucleotides, about 20 nucleotides, about 50 nucleotides, about 100 nucleotides or more. By "portion" or "fragment", as it relates to a nucleic acid molecule, sequence or segment of the invention, when it is linked to other sequences for expression, is meant a sequence having at least 80 nucleotides, more preferably at least 150 nucleotides, and still more preferably at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means at least 6, about 9, preferably 12, more preferably 15, even more preferably at least 20, consecutive nucleotides, e.g., probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the nucleic acid molecules of the invention.

By "resistant" is meant an organism, e.g., a plant or animal, that exhibits substantially no phenotypic changes as a consequence of infection with a

pathogen. By “tolerant” is meant an organism which, although it may exhibit some phenotypic changes as a consequence of infection, does not have a decreased reproductive capacity or substantially altered metabolism.

5 The term “gene” is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for
10 other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

 “Naturally occurring” is used to describe an object that can be found in
15 nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

 A “marker gene” encodes a selectable or screenable trait.

20 “Selectable marker” is a gene whose expression in a cell gives the cell a selective advantage. The selective advantage possessed by the cells transformed with the selectable marker gene may be due to their ability to grow in the presence of a negative selective agent, such as an antibiotic or a herbicide, compared to the growth of non-transformed cells. The selective advantage
25 possessed by the transformed cells, compared to non-transformed cells, may also be due to their enhanced or novel capacity to utilize an added compound as a nutrient, growth factor or energy source. Selectable marker gene also refers to a gene or a combination of genes whose expression in a cell gives the cell both a negative and/or a positive selective advantage.

30 The term “chimeric” refers to any gene or DNA that contains 1) DNA sequences, including regulatory and coding sequences, that are not found

together in nature, or 2) sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or comprise regulatory sequences and coding
5 sequences derived from the same source, but arranged in a manner different from that found in nature.

A “transgene” refers to a gene that has been introduced into the genome by transformation and is stably maintained. Transgenes may include, for example, DNA that is either heterologous or homologous to the DNA of a
10 particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term “endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

15 The terms “protein,” “peptide” and “polypeptide” are used interchangeably herein.

By “variants” is intended substantially similar sequences. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein.
20 Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis which encode the
25 native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have at least 40, 50, 60, to 70%, e.g., preferably 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, e.g., 81%-84%, at least 85%, e.g., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98%,
30 sequence identity to the native (endogenous) nucleotide sequence.

- “DNA shuffling” is a method to introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is
- 5 a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA preferably encodes a variant polypeptide modified with respect to the polypeptide encoded by the template DNA, and may have an altered biological activity with respect to the polypeptide encoded by the template DNA.
- 10 The nucleic acid molecules of the invention can be optimized for enhanced expression in an organism of interest (Wada et al., Nucl Acids Res., 18:2367 (1990). For plants see, for example, EPA035472; WO91/16432; Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324 (1991); and Murray et al., Nucl Acids Res. 17:477 (1989). In this manner, the genes or gene fragments can be
- 15 synthesized utilizing plant-preferred codons. See, for example, Campbell and Gowri, 1990 for a discussion of host-preferred codon usage. Thus, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the gene sequence may be optimized or synthetic. That is, synthetic or partially optimized sequences may also be used.
- 20 Variant nucleotide sequences and proteins also encompass sequences and protein derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different coding sequences can be manipulated to create a new polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a
- 25 population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer, Nature, 370:389 (1994); Cramer et al., Nature Biotech., 15:436 (1997); Moore et al., JMB, 272:336 (1997); Zhang et al., Proc. Natl.
- 30 Acad. Sci. USA, 94:4504 (1997); Cramer et al., Nature, 391:288 (1998); and U.S. Patent Nos. 5,605,793 and 5,837,458.

“Conservatively modified variations” of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences.

5 Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the
10 encoded protein. Such nucleic acid variations are “silent variations” which are one species of “conservatively modified variations.” Every nucleic acid sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the
15 only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each “silent variation” of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

“Recombinant DNA molecule” is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to
20 join together DNA sequences as described, for example, in Sambrook et al., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1989).

The terms “heterologous DNA sequence,” “exogenous DNA segment” or “heterologous nucleic acid,” each refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified
25 from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or
30 homologous to the cell but in a position within the host cell nucleic acid in which

the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

A "microarray" as used herein is a solid support and a plurality of different oligonucleotides attached to the support. Each of the different
5 oligonucleotides is attached to the surface of the solid support in a different defined region, has a different determinable sequence, and is at least six nucleotides in length. Preferably, at least one of the different oligonucleotides is derived from a region of a polynucleotide having a nucleotide sequence selected from SEQ ID NO:46, SEQ ID NO:48 and SEQ ID NO:55, or the complement
10 thereof.

A "homologous" DNA sequence is a DNA sequence that is naturally associated with a host cell into which it is introduced.

"Wild-type" refers to the normal gene, e.g., a gene found in the highest frequency in a particular population, or organism found in nature without any
15 known mutation.

"Genome" refers to the complete genetic material of an organism.

"Vector" is defined to include, inter alia, any plasmid, cosmid, phage or binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform
20 prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g., autonomous replicating plasmid with an origin of replication).

Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host
25 organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g., higher plant, mammalian, yeast or fungal cells).

"Cloning vectors" typically contain one or a small number of restriction endonuclease recognition sites at which foreign DNA sequences can be inserted in a determinable fashion without loss of essential biological function of the
30 vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically

include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

“Expression cassette” as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

Such expression cassettes will comprise the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Such an expression cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

A transcriptional cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source. For expression in plants,

convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau et al., Mol. Gen. Genetics, 262:141 (1991); Proudfoot, Cell, 64:671 (1991); Sanfacon et al., Genes Dev., 5:141 (1991);
5 Mogen et al., Plant Cell, 2:1261 (1990); Munroe et al., Gene, 91:151 (1990); Ballas et al., Nucl. Acids Res., 17:7891 (1989); Joshi et al., Nucl. Acids Res., 15:9827 (1987).

An oligonucleotide corresponding to a nucleic acid molecule of the invention may be about 30 or fewer nucleotides in length (e.g., 9, 12, 15, 18, 20,
10 21 or 24, or any number between 9 and 30). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length may be preferred. Those skilled in the art are well versed in the design of primers for use processes such as PCR. If required, probing can be done with entire restriction fragments of the
15 gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

“Coding sequence” refers to a DNA or RNA sequence that codes for a specific amino acid sequence and excludes the non-coding sequences 5' and 3' to the coding sequence. It may constitute an “uninterrupted coding sequence”,
20 i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions, e.g., as may be found in genomic DNA. An “intron” is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

25 The terms “open reading frame” and “ORF” refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms “initiation codon” and “termination codon” refer to a unit of three adjacent nucleotides (“codon”) in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA
30 translation).

A “functional RNA” refers to an antisense RNA, ribozyme, or other RNA that is not translated.

The term “RNA transcript” refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA” (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

“Regulatory sequences” and “suitable regulatory sequences” each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. As is noted above, the term “suitable regulatory sequences” is not limited to promoters. However, some suitable regulatory sequences useful in the present invention will include, but are not limited to constitutive promoters, tissue-specific promoters, development-specific promoters, inducible promoters and viral promoters.

“5' non-coding sequence” refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency (Turner et al., Mol. Biotech., 3:225 (1995)).

“3' non-coding sequence” refers to nucleotide sequences located 3' (downstream) to a coding sequence and include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting

mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., Plant Cell, 1, 671, 1989.

5 “Promoter” refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. “Promoter” includes a minimal promoter that is a short DNA sequence comprised of a TATA- box and other sequences that serve to
10 specify the site of transcription initiation, to which regulatory elements are added for control of expression. “Promoter” also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the
15 latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either
20 upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also
25 contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

 The “initiation site” is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With
30 respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (i.e. further protein encoding sequences in the

3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as “minimal or core promoters.” In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A “minimal or core promoter” thus consists only of all basal elements needed for transcription initiation, e.g., a TATA box and/or an initiator.

“Constitutive expression” refers to expression using a constitutive or regulated promoter. “Conditional” and “regulated expression” refer to expression controlled by a regulated promoter.

“Operably-linked” refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory DNA sequence is said to be “operably linked to” or “associated with” a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

“Expression” refers to the transcription and/or translation of an endogenous gene or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

“Altered levels” refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

“Overexpression” refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed cells or organisms.

“Antisense inhibition” refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

5 “Co-suppression” and “transwitch” each refer to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar transgene or endogenous genes (U.S. Patent No. 5,231,020).

“Gene silencing” refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the
10 affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes (English et al., Plant Cell, 8:179 (1996). Gene silencing includes virus-induced gene silencing (Ruiz et al., Plant Cell, 10:937 (1998).

“Chromosomally-integrated” refers to the integration of a foreign gene or
15 DNA construct into the host DNA by covalent bonds. Where genes are not “chromosomally integrated” they may be “transiently expressed.” Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette, for example, or as part
20 of another biological system such as a virus.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) “reference sequence”, (b) “comparison window”, (c) “sequence identity”, (d) “percentage of sequence identity”, and (e) “substantial identity”.

25 (a) As used herein, “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, “comparison window” makes reference to a
30 contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or

deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art
5 understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can
10 be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller, CABIOS, 4:11 (1988); the local homology algorithm of Smith et al., Adv. Appl. Math., 2:482 (1981); the homology alignment algorithm of Needleman and Wunsch, JMB, 48:443 (1970); the search-for-similarity-method
15 of Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85:2444 (1988); the algorithm of Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 87:2264 (1990), modified as in Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 90:5873 (1993).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such
20 implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA).
25 Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al., Gene, 73:237 (1988); Higgins et al., CABIOS, 5:151 (1989); Corpet et al., Nucl. Acids Res., 16:10881 (1988); Huang et al., CABIOS, 8:155 (1992); and Pearson et al., Meth. Mol. Biol., 24:307 (1994). The ALIGN program is based on the algorithm of
30 Myers and Miller, *supra*. The BLAST programs of Altschul et al., JMB,

215:403 (1990); Nucl. Acids Res., 25:3389 (1990), are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information

5 (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990,

10 *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and

15 N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-

20 scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993), *supra*). One measure of similarity provided by the BLAST algorithm is the smallest sum probability

25 (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more

30 preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al., 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al.,
5 *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid
10 sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See <http://www.ncbi.nlm.nih.gov>. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide
15 sequences for determination of percent sequence identity to the sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By “equivalent program” is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or
20 amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum
25 correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with
30 similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in

conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity.” Means for making this adjustment are well known to those of skill
5 in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is
10 calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, “percentage of sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison
15 window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions,
20 dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term “substantial identity” of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%,
25 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will
30 recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by

taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

5 Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of
10 about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic
15 code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term “substantial identity” in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%,
20 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of
25 Needleman and Wunsch, 1970, *supra*. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

30 For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence

comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984; $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\%\text{-form}) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m , hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with

>90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, 1993. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous nucleotide sequences that are substantially

identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may results form, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel, Proc. Natl. Acad. Sci. USA, 82:488 (1985); Kunkel et al., Meth. Enzymol., 154:367 (1987); U. S. Patent No. 4,873,192; Walker and Gaastra, Techniques in Mol. Biol. (MacMillan Publishing Co. (1983), and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al., Atlas of Protein Sequence and

Structure (Nat'l. Biomed. Res. Found. 1978). Conservative substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the polypeptides of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired activity. The deletions, insertions, and substitutions of the polypeptide sequence encompassed herein are not expected to produce radical changes in the characteristics of the polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). See also, Creighton, 1984. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

"Germline cells" refer to cells that are destined to be gametes and whose genetic material is heritable.

The word "plant" refers to any plant, particularly to seed plant, and "plant cell" is a structural and physiological unit of the plant, which comprises a cell

wall but may also refer to a protoplast. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ.

5 “Plant tissue” includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

10 The term “altered plant trait” means any phenotypic or genotypic change in a transgenic plant relative to the wild-type or non-transgenic plant host.

The term “transformation” refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as “transgenic” cells, and organisms comprising transgenic cells are
15 referred to as “transgenic organisms”. Examples of methods of transformation of plants and plant cells include *Agrobacterium*-mediated transformation (De Blaere et al., Meth. Enzymol., 143:277 (1987) and particle bombardment technology (Klein et al., Nature, 327:70 (1987); U.S. Patent No. 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known
20 to the skilled artisan (see, for example, Fromm et al., Biotech., 8:833 (1990).

“Transformed,” “transgenic,” and “recombinant” refer to a host cell or organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome generally known in the art and are disclosed in
25 Sambrook et al., 1989, *supra*. See also Innis et al., PCR Protocols, Academic Press (1995); and Gelfand, PCR Strategies, Academic Press (1995); and Innis and Gelfand, PCR Methods Manual, Academic Press (1999). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers,
30 vector-specific primers, partially mismatched primers, and the like. For example, “transformed,” “transformant,” and “transgenic” plants or calli have

been through the transformation process and contain a foreign gene integrated into their chromosome. The term “untransformed” refers to normal plants that have not been through the transformation process.

5 A “transgenic” organism is an organism having one or more cells that contain an expression vector.

“Transiently transformed” refers to cells in which transgenes and foreign DNA have been introduced but not selected for stable maintenance.

“Stably transformed” refers to cells that have been selected and regenerated on a selection media following transformation.

10 “Genetically stable” and “heritable” refer to chromosomally-integrated genetic elements that are stably maintained in the plant and stably inherited by progeny through successive generations.

“Enzyme activity” means herein the ability of an enzyme to catalyze the conversion of a substrate into a product. A substrate for the enzyme comprises
15 the natural substrate of the enzyme but also comprises analogues of the natural substrate which can also be converted by the enzyme into a product or into an analogue of a product. The activity of the enzyme is measured for example by determining the amount of product in the reaction after a certain period of time, or by determining the amount of product in the reaction after a certain period of
20 time, or by determining the amount of substrate remaining in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of an unused co-factor of the reaction remaining in the reaction mixture after a certain period of time or by determining the amount of used co-factor in the reaction mixture after a certain period of
25 time. The activity of the enzyme is also measured by determining the amount of a donor of free energy or energy-rich molecule (e.g., ATP, phosphoenolpyruvate, acetyl phosphate or phosphocreatine) remaining in the reaction mixture after a certain period of time or by determining the amount of a used donor of a free energy or energy-rich molecule (e.g., ADP, pyruvate, acetate or creatine) in the
30 reaction mixture after a certain period of time.

“Fungicide” is a chemical substance used to kill or suppress the growth of fungal cells.

An “inhibitor” is a chemical substance that causes abnormal growth, e.g., by inactivating the enzymatic activity of a protein such as biosynthetic enzyme, receptor, signal transduction protein, structural gene product, or transport protein that is essential to the growth or survival, or alters the virulence or pathogenicity, of the fungus. In the context of the instant invention, an inhibitor is a chemical substance that alters the activity encoded by any one of SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:56 or their orthologs.

“Isogenic” fungi are genetically identical, except that they may differ by the presence or absence of a heterologous DNA sequence.

A “substrate” is the molecule that an enzyme naturally recognizes and converts to a product in the biochemical pathway in which the enzyme naturally carries out its function, or is a modified version of the molecule, which is also recognized by the enzyme and is converted by the enzyme to a product in an enzymatic reaction similar to the naturally-occurring reaction.

“Tolerance” as used herein is the ability of an organism, e.g., a fungus, to continue essentially normal growth or function when exposed to an inhibitor or fungicide in an amount sufficient to suppress the normal growth or function of native, unmodified fungi.

The Nucleic Acid Molecules of the Invention and Uses Thereof

The involvement of peptide synthetase genes in fungal pathogenesis to plants has been genetically tested only in two previous studies. In *C. carbonum*, disruption of both copies of the *HTS1* gene, which encodes HC-toxin synthetase, caused loss of ability to make HC-toxin and the fungus became nonpathogenic on HC-toxin sensitive corn plants (Panaccione et al, PNAS, 89, 6590, 1992), indicating that the HC-toxin synthetase gene is a pathogenicity determinant. In *Fusarium avenaceum*, the enniatin-nonproducing transformants were obtained by disruption of enniatin synthetase encoding gene (*esyn1*) and these transformants displayed significantly reduced virulence in a potato tuber tissue assay

(Herrmann et al., 1996) indicating that enniatin synthetase gene is a virulence factor in pathogenesis by the fungus. In these two pathosystems, only one fungal secondary metabolite (the peptide toxin) was studied. In contrast, the polyketide T-toxin has been well studied in *C. heterostrophs* and has been confirmed to be a host-specific virulence factor (Yoder and Turgeon, 1996; Yoder et al., 1997, *supra*) and this study demonstrated that a second secondary metabolite, the hypothetical CPS1 toxin is also involved in pathogenesis by the fungus. Unlike the T-toxin biosynthetic genes such as *PKS1* and *DEC1* that are found only in race T (Yang et al., 1996, *supra*; Rose et al., 1996, *supra*), *CPS1* is found in both race O and race T. Disruption of *CPS1* in either race causes dramatically reduced fungal virulence as tested on N-cytoplasm corn. This result suggests that CPS1 toxin could be the same as the “race O” toxin proposed previously (Yoder, 1981). However, as disclosed herein, CPS1 is a CoA ligase.

Interestingly, a *Tox*⁺, *cps1*⁻ mutant also show reduced virulence on T-cytoplasm corn although it produced the same amount of T-toxin as wild type race T. This is unusual because the interaction between T-toxin and the T-corn-unique URF13 protein is highly specific; the same outcomes should be expected if two strains that produce the same amount of T-toxin attack the same host, T-corn. The most likely explanation for this result is that the fungal growth *in planta* has been inhibited by the host plant and the poor growth results in reduced T-toxin production which is normal when the fungus is grown in culture. Reduced virulence on T-cytoplasm corn is due to the reduced T-toxin production as that seen in leaky *Tox*⁻ mutants. This inhibition of growth could be due to the failure of suppression of the host defense mechanism by the fungus, which is mediated by the *CPS1* controlled peptide toxin. A *cps1*⁻ mutant that fails to produce this “suppressor” could not be able to colonize plant tissues as vigorously as wild type does, resulting in the reduced ability to cause disease as indicated by the smaller lesion phenotype. If this turns out to be the case, *CPS1* should be considered as a general virulence factor as proposed for enniatin.

It is possible that *cps1*⁻ mutants are still be able to produce a certain amount of CPS1 toxin. One probability is the gene has not been completely

inactivated by insertional mutagenesis or targeted disruption. The original REMI insertion occurred at core sequence 1 of CPS1A, a region that might be not critical (function of core 1 is unknown). The second targeted site is located between cores 1 and 2 of CPS1B and the third is located between cores 2 and 3 of the same module. All three insertions do not disrupt critical motifs. On the other hand, *CPSI* contains a number of in-frame start codons and some of them are located immediately downstream of these insertion sites. It is possible that each of these disruptions actually resulted in two subtranscripts, one is transcribed normally from the start codon of *CPSI* and stops at the insertion site and second is transcribed near one of these in-frame ATGs downstream of the insertion site and stops at the end of *CPSI*. Both transcripts could give a truncated protein that still has enzymatic activities. But these separate enzymes might have affinities for their substrates lower than that of holoenzyme. The reduced production of CPS1 toxin might be due to the CPS1 holoenzyme having been split into two fractions by the vector insertion and the resulting truncated proteins being much less active than the original polypeptide. This hypothesis can be tested by construction a *C. heterostrophus* strain in which the entire *CPSI* encoding sequence has been deleted.

The second possibility is the existence of multiple copies of *CPSI* in the genome. Previous studies have demonstrated that the gene encoding HC-toxin synthetase (*HTSI*) is duplicated in the genome and both copies (*HTSI-1* and *HTSI-2*) are 270 kb apart in most Tox2⁺ isolates of *C. carbonum* (Ahn and Walton, Plant Cell, 8, 887, 1996). Disruption of either copy reduced HTS1 activity but did not affect HC-toxin production; when both copies were

25

disrupted, HC-toxin production was abolished (Panaccione et al, 1992, *supra*). But in contrast to the case of *HTSI*, gel blot analysis does not indicate the presence of a second copy of *CPSI* and disruption of *CPSI* does affect the production of the putative toxin. It is unlikely that two genes with similar organization are in the genome. An alternative postulation is that there may be a second gene which encodes a protein with the same enzyme activity as *CPSI* but does not have significant sequence homology to *CPSI*. This hypothesis is hard to test unless this gene is clustered with *CPSI* and can be recovered by chromosome walking.

10 Pathogenesis by *C. heterostrophus* to corn involves at least two secondary metabolites: the T-toxin, a host specific factor which determines high virulence on a particular host, T-corn and the hypothetical *CPSI* toxin, a general factor (either virulence or pathogenicity factor) which contributes to basic mechanisms underlying the disease establishment by the fungus in common host plants.

15 By genomic DNA hybridization, *C. heterostrophus CPSI* homologs were found in 16 additional fungal species belonging to 5 genera. Hybridization signals for some were as strong as the *C. heterostrophus* gene, indicating that *CPSI* is highly conserved among these fungi. This conservation appears to match the taxonomic relationships between these species. *Cochliobolus* (anamorph *Bipolaris*) and *Setosphaeria* (anamorph *Exserohilum*) are closely related genera.

20 Two species, *C. victoriae* and *C. carbonum*, which are able to cross to each other and thus may not be different species (Scheffer et al., 1967; Yoder et al., 1989), showed the same hybridization pattern to *CPSI*. *B. sacchari*, the closest asexual relative of *C. heterostrophus*, hybridized to two *HindIII* fragments that were only seen in *C. heterostrophus* itself, but all other species gave only one distinct polymorphic band. Phylogenetic analyses using the internal transcribed spacer (ITS) sequences and fragments of the *GPD* (vanWert and Yoder, 1992) and *MAT* genes (Turgeon et al., Mol. Gen. Genet., 238, 270, 30 1993) also put *C. victoriae/C. carbonum* and *C. heterostrophus/B. sacchari*

closest to each other (Turgeon and Berbee, 1997). These results might imply that *CPSI* has coevolved with these genes.

The genera *Cochliobolus* and *Setosphaeria* include many plant pathogenic species that are commonly associated with leaf spots or blights, mainly on cultivated cereals and wild grasses (Sivanesan, 1987; Alcorn, 1988). This group of phytopathogenic fungi includes both mild pathogens and severe pathogens that often produce host-specific toxins (Yoder, 1980, *supra*). One of the essential questions is whether or not the various diseases on diverse host plants caused by these fungi involve common factors or depend only on individual specific factors, such as host-specific toxins.

Previous studies have shown that host-specific toxins can be critical factors for determining either virulence or host-range, but they do not account for general pathogenicity since they are produced only by certain isolates in the species and the corresponding biosynthetic genes are found only in these toxin-producing isolates (Yoder et al., 1997, *supra*). In contrast, *CPSI* homologs are found in all *Cochliobolus* and *Setosphaeria* species tested so far, suggesting they are a common factor shared by this group. Disruption of the *CPSI* homolog in the oat pathogen *C. victoriae* caused dramatically reduced virulence to victorin-susceptible oats although the transformants produced wild type levels of victorin. This result is similar to that with *C. heterostrophus* race T, in which *cpsI*⁻ disruptants still produced wild type levels of T-toxin but showed reduced virulence on T-cytoplasm corn. These results argue strongly that host-specific toxins alone are not sufficient in determining the ultimate outcome of fungus/plant interactions and suggest that the establishment of disease by these fungi also requires CPS1, which might control a pathway for general pathogenicity.

In the early 1990s, studies on pathogenesis by uropathogenic *E. coli* led to the identification of pathogenicity gene clusters, termed “pathogenicity islands” (Hecker et al., 1990; Blum et al., 1994). Subsequently, similar gene clusters were identified in additional animal or human bacterial pathogens, including *Yersinia pestis*, *Helicobacter pylori* and *Salmonella typhimurium*.

These islands often contain genes for production of toxins or genes encoding proteins that are capable of interacting with host defense factors or required for type III secretion systems that deliver virulence proteins into host cells. Usually, they are found only in pathogenic strains (or species); in rare cases, they occur in nonpathogenic strains of the same species or related species (Hacker et al., Mol. Microbiol., 23, 1089, 1997).

In phytopathogenic bacteria, *hrp* gene clusters have been referred to as “pathogenicity islands” because they have several features in common with “pathogenicity islands” in animal pathogenic bacteria, i.e., they are found only in pathogenic species (required for plant pathogenicity) and contain highly conserved genes (*hrc* genes) defining the type III protein secretion system (Alfano and Collmer, 1996; Barinaga, 1996).

In plant pathogenic fungi, genes or gene clusters with characteristics of “pathogenicity islands” have been identified from certain species, i.e., in *Nectria haematococca*, the *PDA* genes for detoxifying the pea phytoalexin and other pea pathogenicity genes (*PEP*) are located on dispensable chromosomes that are found in all isolates pathogenic to pea but usually absent in all nonpathogenic isolates (VanEtten et al., Antonie Van Leeuwenhoek, 65, 263, 1994; Liu et al., 1997, *supra*). In the genus *Cochliobolus*, the *Tox2* gene cluster controlling the biosynthesis of HC-toxin is found only in *C. carbonum* race 1 (pathogenic to *hm1hm1* corn) and the *Tox1* genes controlling T-toxin production are found only in *C. heterostrophus* race T (highly virulent on T-cytoplasm corn); all other races of the same species and all other fungal species tested so far lack these *Tox* genes (Ahn and Walton, 1996, *supra*; Yang et al., 1996, *supra*; Yoder et al., 1997, *supra*).

CPS1 differs in two important ways compared to these fungal “pathogenicity islands”. First, it is highly conserved among several phytopathogenic *Cochliobolus* species and relatives. Second, like certain bacterial “pathogenicity islands”, *CPS1* also has homologs in “nonpathogenic” species. *C. homomorphus* and *C. dactyloctenii*, neither of which causes disease on plants, hybridized strongly to *CPS1*. This may reflect genetic changes in the

“pathogenicity island” that resulted in loss of pathogenicity. In the bacterial genus *Listeria*, which includes several human or animal pathogenic species harboring highly conserved “pathogenicity islands”, the “pathogenicity island” homolog in the nonpathogenic species (*L. seeligeri*) was found to be “silent” due to a mutation that occurred in the promoter region of a critical regulatory gene in the cluster (Hacker et al., 1997, *supra*). These features suggest that the *CPSI* gene cluster and homologs could define a new group of fungal “pathogenicity islands”.

It is known that the evolution of pathogenicity involves two major processes. A pathogenic microorganism could originate from nonpathogenic progenitors by slow modifications (such as point mutations and genetic recombination) of genes that were adapted for parasitic growth on hosts or by the integration of large fragments of “alien” DNA into the genome that enable the recipient to attack particular hosts (gene horizontal transfer). The latter can occur in the recent or distant evolutionary past. Subsequent vertical transmission in the lineage (if the transferred gene is stable in the recipient genome) would result in the preserve of the gene in all species that diverged after the acquisition of the gene(s) (Scheffer, 1991; Arber, *Gene*, 135, 49, 1993; Krishnapillai, 1996; Burdon and Silk, 1997).

In the past few years, substantial evidence has become available that supports the hypothesis of gene horizontal transfer. All “pathogenicity islands” in animal pathogenic bacteria are believed to have been acquired by a horizontal transfer event (recent or past) because they usually differ in G+C content from the recipient genome and have transposable elements at the boundaries of the gene clusters (Hacker et al., 1997, *supra*). The *hrp* “pathogenicity islands” do not show a significant difference in G+C content or association with transposable elements, but they are also believed to have arisen similarly because *hrc* genes in these “pathogenicity islands” show high similarity to genes defining the type III protein secretion system found in animal pathogenic bacteria as mentioned above (Alfano and Collmer, 1996; and Barinaga, 1996).

Although *CPSI* itself has several typical fungal introns and a G+C content (51.5%) similar to most known fungal genes, genomic regions (about 1.5 kb) flanking the gene have higher G+C content (>60%). Several short G+C-rich regions are also found in the gene cluster; one of the open reading frames (ORF10) has a 63.6 % G+C content. Compared to those filamentous fungal genomes characterized so far, including *N. crassa*, *A. nidulans*, *U. maydis* (all have G+C content 51-54%, see Karlin and Mrázek, PNAS, 94, 10227, 1997), the genomic region around *CPSI* is unusual. This might suggest that the gene cluster harboring *CPSI* came from a bacterial source (since most bacterial genes are known to have a high G+C content), but has evolved into a fungal version.

Based on these data, *CPSI* homologs may have a common ancestral gene which was acquired from a bacterial species *via* horizontal transfer and then maintained by the fungal genome *via* vertical transmission in closely related lineages.

In the evolution process, the genus *Cochliobolus* could also have inherited a second gene (*X*) controlling the ability to take up foreign DNA, by which its ancestor took the “alien” *CPSI*. As a result, this group of fungi is able to keep trapping genes from other organisms by additional “horizontal transfers” and giving rise to new races or even new species characterized by the ability to produce unique pathogenesis factors. The direct support for this hypothesis is that both the *Tox2* locus of *C. carbonum* and the *Tox1* locus of *C. heterostrophus* are associated with large fragments of “alien” DNA (A+T-rich and highly repeated) and the same could also be true for *Tox3* controlling victorin production by *C. victoriae*, although there is yet no direct experimental evidence (Ahn and Walton, 1996, *supra*; Yang et al., 1996, *supra*; Yoder et al., 1997, *supra*). In contrast to *CPSI*, these gene transfers must have occurred in the recent evolutionary past because both *Tox1* and *Tox2* loci are found only in specific isolates in the species, e.g., the acquisition of *Tox1* genes probably occurred as recently as the 1960s when race T was first identified in the field (Yoder et al., 1997, *supra*).

There are other possibilities for the evolution of *CPSI*. First, each genus mentioned above could have acquired *CPSI* independently after divergence of the lineage. But this seems less likely because this would need to happen at the same time and involve the same donor organism if the fact that the homologs
5 detected in *Cochliobolus* and *Setosphaeria* gave similar hybridization signal intensity is considered. Second, the horizontal transfer of *CPSI* could have occurred at earlier time periods such as before the divergence of Pleosporales or even the Ascomycotina. To test these hypotheses, detection of *CPSI* homologs in *Pyrenophora*, *Pleospora* and other genera must be done by either genomic
10 DNA hybridization or PCR. Based on the facts discussed here, it is not unreasonable to predict that additional *CPSI* homologs will be found in other fungal species. Further investigation could provide an direct entry point for understanding the evolution of fungal pathogenesis to plants.

The *C. heterostrophus CPSI* gene was cloned by identification of
15 genomic DNA fragments recovered from the tagged site in a mutant generated using REMI insertional mutagenesis. Characterization of two overlapping cosmid clones in this study has proved that no deletions or chromosome rearrangements are associated with the gene tagging event, because both cosmids carry the same fragment which span the REMI insertion site and the nucleotide
20 sequence in this region is the same as that of recovered genomic DNA from the tagged site. This undoubtedly clarifies the identity of *CPSI*, which is the major biosynthetic gene. Mapping and sequencing of the two cosmids extended the sequence by 27.4 kb from the previously cloned fragment, leading to the characterization of 38.7 kb of contiguous genomic DNA, the largest genomic
25 region analyzed so far in *C. heterostrophus*. In addition to *CPSI* and *TESI*, sequence analysis of this region revealed at least 11 open reading frames; three of them, designated as *DBZ1*, *CAT1* and *DEC2*, respectively, apparently encode functional proteins. The tight linkage of these genes suggests that they may be involved in the same pathway.

30 In filamentous fungi, in some cases, genes in pathways for biosynthesis of secondary metabolites are dispersed on different chromosomes, e.g., the

cephalosporin C pathway genes in *Acremonium chrysogenum* (Mathison et al., Curr. Genet., 23, 33, 1993) and the melanin pathway genes in *Colletotrichum lagenarium* (Kubo et al., Appl. Environ. Microbiol., 62, 4340, 1996). In other cases, tightly linked genes are usually found to be functionally related to a common pathway. This clustering organization has been exemplified by the sterigmatocystin pathway genes of *Aspergillus nidulans*, in which coordinately regulated transcripts are found in a 60 kb genomic region (Brown et al., 1996) and the trichothecene pathway genes of *Fusarium sporotrichioides*, in which 9 genes are clustered in a 25 kb region and 8 of them have been shown to be required for the pathway function (Hohn et al., Mol. Gen. Genet., 248, 95, 1995). The genes involved in biosynthesis of certain fungal peptides are also found as clusters. The tight linkage between *CPSI* and these additional genes might reveal the presence of a novel secondary metabolite pathway in *C. heterostrophus*. In this pathway, *CPSI* is the major structural gene since it encodes a large multifunctional enzyme with all catalytic activities required for synthesis of a secondary metabolite, presumably a peptide phytotoxin; other genes may carry out different functions required for coordinate operation of the pathway, such as regulation, posttranslational modification or substrate processing as discussed below.

Both functional and structural analyses strongly support the hypothesis that the *CPSI* gene cluster controls a novel biosynthetic pathway. Pathway genes have been studied only in a few filamentous fungi mainly for industrial purposes (Keller et al., J. Ind. Microbiol. Biotechnol., 19, 305, 1997). For plant pathogenic fungi, little is known about pathway genes for fungal pathogenesis. In *C. heterostrophus*, recent cloning of two *ToxI* genes *PKSI* (Yang et al., 1996, *supra*) and *DEC1* (Rose et al., 1996, *supra*) have contributed to a breakthrough in understanding the molecular mechanism for biosynthesis of T-toxin, a virulence determinant in the fungus/corn interaction. But further identification of related pathway genes has been unsuccessful because the two genes are located on different chromosomes and each is embedded in A+T-rich DNA

(Yoder et al., 1997, *supra*). In contrast, the *CPSI* cluster provides a good opportunity to explore a pathogenesis pathway.

First, it resides in a “normal” sequence region. G+C content of a 50-55% is found in most of the cloned sequences and no A+T-rich DNA is associated with either end of the cloned region. This would facilitate cloning of additional pathway genes by further chromosome walking, by screening of cosmid libraries or the targeted integration and plasmid rescue. Second, it contains a regulatory gene (*DBZI*) which is presumably linked to a signal transduction pathway. Isolation of genes that interact with *DBZI* could reveal novel factors mediating the molecular communication between fungal pathogen and the host plant. Further characterization of *DBZI* (along with position-specific disruption or deletion) would be also helpful in determining the limit of the gene cluster, because tightly linked genes involved in a common pathway are often coordinately regulated by the same regulatory factor (Keller et al., 1997, *supra*). Finally, *CPSI* genes are found in both race T and race O, and its homologs are also found in other *Cochliobolus* species. Presence of high G+C content may imply that these genes evolved from a bacterial ancestor and the conservation in these fungi may correlate with the phytopathogenic function of the gene products encoded by the *CPSI* cluster. Further investigation of this cluster should provide insights into the evolution of general pathogenicity factors among this group of fungi.

Ferric reductases are a group of enzymes found in bacteria, fungi, plants and animals that are responsible for reduction of ferric iron to ferrous iron, an absorptive form used by the organism. They have been well studied in *S. cerevisiae*, *C. albicans* and *H. capsulatum* and the like. The yeast FER1 has been expressed in tobacco (Oki et al., 1999).

Previous studies have shown that FER genes could be important pathogenic determinants. Timmerman and Woods have proposed that in *H. capsulatum* FER could play critical roles in the acquisition of iron in three different ways: from inorganic or organic ferric salts, from host Fe(III) binding

proteins (transferrin and the like), and from siderophores produced by the fungus itself (to reduce and release the iron chelated by the siderophore molecules).

On the other hand, iron sequestration in response to microbial infection has been demonstrated to be a host defense mechanism. The infection-related iron acquisition system in the pathogen can be considered to be an important mechanism against host defense and for a successful colonization by the pathogen in the host cells. This could be a general mechanism for all pathogenic fungi.

CPS1 does encode a peptide synthetase which is responsible for biosynthesis of a novel siderophore with unusual amino acid, hydroxyl acid and architecture, which is why CPS1 does not show similarity to common NRPSs. The CPS1 siderophore can compete with the host for iron acquisition when the fungus enters its host cells where the iron is limited due to host sequestration. In particular, for root pathogens such as *C. victoriae*, sequestration may be stronger in the root surface. This could explain why the *cps1* mutant showed drastically reduced virulence. The FER1 could be required to release iron from the CPS1 siderophore which explains its location near the *CPS1* gene. Moreover, fungal strains could be cultured in iron-limiting conditions because CPS1, and likely other genes in the cluster maybe turned on only during conditions of iron depletion.

In a preferred embodiment, the polypeptides, including those having substantially similar activities to SEQ ID NO:47, SEQ ID NO:49, or SEQ ID NO:56 are encoded by nucleotide sequences derived from fungi, preferably from pathogenic fungi, desirably identical or substantially similar to the nucleotide sequences set forth in SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55 or the complement thereof.

In another preferred embodiment, the present invention describes a method for identifying agents having the ability to inhibit or reduce the activity of any one or more of SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56 in fungi. Preferably, a transgenic "knockout" fungus and/or fungal cell, is obtained which preferably is stably transformed, which comprises a deletion in any of

SEQ ID NO:46, SEQ ID NO:48 or SEQ ID NO:55. Thus, in one embodiment, the gene product encoded by the nucleotide sequence is not expressed, or has reduced or aberrant expression. In another embodiment, the transgenic fungus or cell comprises the corresponding non-deleted sequences linked to a promoter to
5 yield a gene product which is overexpressed. An agent is then contacted with the transgenic fungus and/or cell, and the growth development, virulence or pathogenicity of the transgenic fungus and/or cell is determined relative to the growth, development, or pathogenicity, of the corresponding transgenic fungus and/or cell to which the agent was not applied; or to the corresponding non-
10 transgenic fungus and/or cell.

The present invention generally relates to an isolated nucleic acid molecule from a fungal pathogen encoding a CPS1 peptide synthetase, an iron reductase or a permease/MFS transporter. In a preferred embodiment, a DNA molecule has a nucleotide sequence which hybridizes to a DNA molecule having
15 a sequence corresponding to SEQ ID NO:46, SEQ ID NO:48 or SEQ ID NO:55. Other DNA molecules of the present invention include DNA molecules that have a sequence which is greater than 65% identical to the nucleotide sequence of SEQ ID NO:46, SEQ ID NO: 48 or SEQ ID NO:55. Nucleotide sequence similarity is determined by the BLAST program with the default parameters
20 (Altschul et al., "Basic Local Alignment Search Tool," *J. Mol. Biol.*, 215:403 (1990). Preferred sequences include those DNA molecules which will hybridize to a nucleic acid molecule having the sequence of SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55 or the complement thereof. Preferably, the DNA molecules hybridize to SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:55, or its
25 complement under low or moderate, or stringent conditions.

Other proteins or polypeptides of the present invention include polypeptides having an amino acid sequence which has at least 75% similarity to the amino acid sequence of SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56. In a preferred embodiment of the invention, the protein or polypeptide will have
30 at least 90% similarity with SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56.

In addition, the nucleic acid molecules of the invention may be modified, adapted, and optimized in such a manner that, when transferred into an appropriate host cell, the modified polynucleotide confers an altered phenotype brought about by the polypeptide encoded by the modified sequence. One
5 advantage of this method is that it can be used to rapidly evolve any protein without knowledge of its structure. Peptide synthetase, iron reductase and/or permease/MFS transporter polynucleotides can be altered using sequence-shuffling methods as described by WO 00/28008 and references therein. Peptide synthetases of the invention can be recombined with other peptide synthetases,
10 iron reductases and/or permeases/MFS transporters to generate peptide synthetases, iron reductases and/or permeases/MFS transporters of desired and/or novel specificity and/or activity, and thus generate desired and/or novel non-encoded peptide products. Such novel peptide synthetases, iron reductases and/or permeases/MFS transporters would have at least one active domain or
15 other desired property-imparting domain (e.g., binding, enzymatic activity, specificity determining).

Briefly, sequences or fragments of sequences are shuffled by various recombinatorial methods, the shuffled polynucleotide is introduced into a suitable host for expression, the resulting phenotype is measured and the
20 modified phenotype is compared with the phenotype produced by unmodified sequence. Here, "phenotype" refers to the trait of interest and may include measuring the amount, conformation, composition, or enzymatic activity of the polypeptide encoded, if the sequence shuffling is being performed, to modify a single protein. Phenotype may also be assessed by measuring the effect of
25 expression of the modified peptide synthetase, iron reductase and/or permease/MFS transporter polynucleotide on expression of other genes, on cellular processes such as respiration or glycolysis, on tissue-level processes such as cell shape and size, and on organismal traits such as pathogenicity and/or virulence. Sequence-shuffled peptide synthetase polynucleotides producing a
30 desirable phenotype are then selected, further modified, and the resulting phenotype is measured. The shuffling and selection process is performed

iteratively until sequence shuffled polynucleotides encoding at least one polypeptide producing the desired phenotype is obtained, or until optimization of the trait of interest has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection. Alternately, multiple rounds of recombination of peptide synthetase sequences may be performed prior to any selection step, with the aim of increasing the diversity of resulting populations nucleic acids prior to selection.

At least five general classes of recombination methods may be applied to peptide synthetase, iron reductase and/or permease/MFS transporter polynucleotides. First, the nucleic acids of peptide synthetase, iron reductase and/or permease/MFS transporter polynucleotides can be recombined *in vitro* by any of a variety of techniques including DNase digestion of polynucleotides followed by ligation and/or PCR reassembly of the polynucleotides. Second, polynucleotides can be recursively recombined *in vivo*, for example by allowing recombination to occur between an introduced peptide synthetase, iron reductase and/or permease/MFS transporter polynucleotide and homologous sequences in a cell. Third, whole cell genome recombination methods can be used in which whole genomes of cells are recombined, optionally including spiking the genomic (nuclear and/or plastid) recombination mixtures with the peptide synthetase, iron reductase and/or permease/MFS transporter sequences of interest. Fourth, synthetic recombination methods can be used, in which oligonucleotides corresponding to different homologs of the peptide synthetase, iron reductase and/or permease/MFS transporter sequence are synthesized and reassembled in PCR or ligation reactions which also include oligonucleotides which correspond to more than one allelic variant, thereby generating new recombined polynucleotides. Fifth, *in silico* methods of recombination can be carried out in which genetic algorithms are used in a computer to recombine sequence strings which correspond to homologs of the peptide synthetase sequences of interest. The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined sequences. Such synthesis could proceed by oligonucleotide

synthesis and gene reassembly techniques. Any of the preceding general recombination formats can be practiced reiteratively to generate a more diverse set of recombinant nucleic acids.

5 The ever-increasing quantity and quality of data being accumulated not only about gene sequence, structure and function, but also about gene expression patterns and proteins interactions on genomic scales, makes it no longer feasible to deal with genetic data on an item-by-item basis but instead, necessary to create new ways of discovering biological information by *in silico* data mining. "Data mining" as used herein, refers to exploration and analysis of large quantities of data, by automatic and semi-automatic means, in order to discover meaningful patterns and rules. Data mining is applied to molecular sequence and structure data, gene expression and other high-throughput data, and to existing knowledge in the scientific literature, including making meaningful connections between different forms of knowledge and data.

15 A variety of data mining tools can be applied using the peptide synthetase, iron reductase and/or permease/MFS transporter sequences of the present invention. A method appropriate for use in sequence databases which contain long stretches of data known as long-pattern data sets, is that disclosed in U.S. Patent No. 6,138,117, which uses a look-ahead scheme for quickly identifying long patterns that is not limited to the initialization phase, an heuristic item-ordering policy for tightly focusing the search, and a support-lower-bounding scheme that is also applicable to other algorithms. Recursive partitioning is useful to elucidate structure-activity relations and to guide decision-making for high-throughput screening of compounds for their effects on peptide synthetase polypeptides, for example as described by Hertzog et al. (J. Pharmacol Toxicol Methods 42:207 (1999)) for sequential screening of G-protein-coupled receptors. The peptide synthetase, iron reductase and/or permease/MFS transporter sequences of the present invention may be applied to digital differential display (DDD) to analyze differential expression and create an electronic expression profile for a variety of physiological conditions. Peptide synthetase, iron reductase and/or permease/MFS transporter sequence data can

be analyzed to predict protein domains using the BLAST algorithm. Higher-order correlations among peptide synthetase, iron reductase and/or permease/MFS transporter proteins may be predicted by using peptide synthetase protein sequence data to compare sets of sequence-distant sites displaying high mutual information which may bespeak important structural or functional features, a methodology that overcomes the limitations of previous methods which examined only single-residue features or pairwise interactions. (Steeg et al., Pac Symp Biocomput 1998:573 (1998)).

Peptide synthetase, iron reductase and/or permease/MFS transporter polypeptide sequences having structures expressed in a computer-readable form can be evaluated for function using functional site descriptors (FSDs) for a biomolecule functional site having a specific biological function, as described in the publication WO 00/11206. FSDs can be used to identify or screen for a novel function in one or more peptide synthetase, iron reductase and/or permease/MFS transporter polypeptides, to confirm a previously identified or suspected function of a protein, to evaluation the effects of sequence shuffling on protein function, or to provide further information about a specific functional site in a peptide synthetase, iron reductase and/or permease/MFS transporter polypeptide.

FSDs are geometric representations of protein functional sites, typically defining spatial configurations of functional sites by providing a three-dimensional (3D) representation of a protein functional site. Preferred functional sites represented by FSDs include a ligand binding domain, an ion or cofactor binding site, a site or domain for protein-protein interaction, or an enzymatic active site. An FSD typically comprises a set of geometric constraints for one or more atoms in each of two or more amino acid residues comprising a function site of a protein. Geometric constraints of an FSD may comprise an atomic position specified by a set of 3D coordinates, an interatomic distance, an interatomic bond angle, or conformational constraints imposed by residues at a site or by secondary structure such as a zinc finger, leucine zipper, helix, or α strand, where these constraints may be expressed either as fixed coordinates or

ranges. Libraries of FSDs can comprise at least two FSDs for at least one of the biological functions represented by the library.

FSDs are used to probe protein structures to determine if such structures contain the functional sites described by the corresponding FSDs. Peptide
5 synthetase, iron reductase and/or permease/MFS transporter polypeptides to be screened can comprise an unmodified sequence selected from SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56, or a modified form derived from random or directed sequence shuffling as previously described. Typically, functional screening methods comprise applying a FSD to a structure of a peptide
10 synthetase, iron reductase and/or permease/MFS transporter polypeptide, where the structure may be determined by x-ray crystallography, nuclear magnetic resonance, by a computer “*ab initio*” folding program a homology program, or a “threading” program, and expressed in a computer-readable form.

The function of a peptide synthetase, iron reductase and/or
15 permease/MFS transporter polypeptide whose structure is expressed in computer-readable form can be screened by applying an FSD to the structure of a peptide synthetase, iron reductase and/or permease/MFS transporter polypeptide and determining whether the peptide synthetase, iron reductase and/or permease/MFS transporter polypeptide structure matches, or satisfies, the
20 constraints of the FSD. Libraries of FSDs can be used to probe for or evaluate the activity or function associated with the FSD in one or more protein structures.

The DNA molecule encoding the CPSI, iron reductase polypeptide and/or permease/MFS transporter of the present invention can be incorporated in cells
25 using conventional recombinant DNA technology. Generally, this involves inserting the DNA molecule into an expression system to which the DNA molecule is heterologous (i.e., not normally present). The heterologous DNA molecule is inserted into the expression system or vector in proper sense orientation and correct reading frame. The vector contains the necessary
30 elements for the transcription and translation of the inserted protein-coding sequences. U.S. Patent No. 4,237,224, describes the production of expression

systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase. These recombinant plasmids are then introduced by means of transformation and replicated in unicellular cultures including prokaryotic organisms and eukaryotic cells grown in culture. Recombinant
5 genes may also be introduced into viruses, such as vaccinia virus. Recombinant viruses can be generated by transfection of plasmids into cells infected with virus.

Suitable vectors include, but are not limited to, the following viral vectors such as lambda vector system gt11, gtWEST.B, Charon 4, and plasmid
10 vectors such as pBR22, pBR325, pACYC177, pACYC184, pUC8, pUC9, pUC18, pUC19, pLG339, pR290, pKC37, pKC101, SV40, pBluescript II SK +/- or KS +/- (see "Stratagene Cloning Systems" Catalog (1993) from Stratagene, La Jolla, Calif), pQE, pIH821, pGEX, pET series (see Studier et. al., "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes," Gene Expression
15 Technology, vol.185 (1990)), and any derivatives thereof. Suitable vectors are continually being developed and identified. Recombinant molecules can be introduced into cells via transformation, transduction, conjugation, mobilization, or electroporation. The DNA sequences are cloned into the vector using standard cloning procedures in the art, as described by Maniatis et al. or
20 Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Springs Laboratory, Cold Springs Harbor, New York (1982 or 1989, respectively).

A variety of host-vector systems may be utilized to express the protein-encoding sequence(s). Primarily, the vector system must be compatible with the host cell used. Host-vector systems include but are not limited to the following:
25 bacteria transformed with bacteriophage DNA, plasmid DNA) or cosmid DNA; microorganisms such as yeast containing yeast vectors; mammalian cell systems infected with virus (e.g., vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (e.g., baculovirus); and plant cells infected by bacteria or transformed via particle bombardment (i.e., biolistics). The expression elements
30 of these vectors vary in their strength and specificities. Depending upon the host-vector system utilized, any one of a number of suitable transcription and

translation elements can be used. Different genetic signals and processing events control many levels of gene expression (e.g., DNA transcription and messenger RNA, "mRNA" translation). Transcription of DNA is dependent upon the presence of a promoter which is a DNA sequence that directs the binding of RNA polymerase and thereby promotes mRNA synthesis. The DNA sequences of eukaryotic promoters differ from those of prokaryotic promoters. Furthermore, eukaryotic promoters and accompanying genetic signals may not be recognized in or may not function in a prokaryotic system, and, further, prokaryotic promoters are not recognized and do not function in eukaryotic cells.

Similarly, translation of DNA in procaryotes depends upon the presence of the proper prokaryotic signals which differ from those of eukaryotes. Efficient translation of DNA in procaryotes requires a ribosome binding site called the Shine-Dalgarno ("SD") sequence on the mRNA. This sequence is a short nucleotide sequence of mRNA that is located before the start codon, usually AUG, which encodes the amino-terminal methionine of the protein. The SD sequences are complementary to the 3'-end of the 16S, rRNA (ribosomal RNA) and probably promote binding of mRNA to ribosomes by duplexing with the rRNA to allow correct positioning of the ribosome. For a review on maximizing gene expression, see Koberts and Lauer, Methods in Enzymology 68:473 (1979).

Promoters vary in their "strength" (i.e., their ability to promote transcription). For the purposes of expressing a cloned gene, it is desirable to use strong promoters in order to obtain a high level of transcription and, hence, expression of the gene. Depending upon the host cell system utilized, any one of a number of suitable promoters may be used. For instance, when cloning in *E. coli*; its bacteriophages, or plasmids, promoters such as the phage promoter, *lac* promoter, *trp* promoter, *recA* promoter, ribosomal RNA promoter, the PR and PL promoters of coliphage lambda and others, including but not limited, to *lacUV5*, *ompF*, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lacUV5* (tac) promoter or other *E. coli* promoters produced by recombinant DNA or other synthetic DNA techniques may be used to provide for transcription of the insert

gene. Bacterial host cell strains and expression vectors may be chosen which inhibit the action of the promoter unless specifically induced. In certain operons, the addition of specific inducers is necessary for efficient transcription of the inserted DNA. For example, the lac operon is induced by the addition of lactose or IPTG (isopropylthiobeta-D-galactoside). A variety of other operons, such as *trp*, *pro*, etc., are under different controls. Specific initiation signals are also required for efficient gene transcription and translation in prokaryotic cells. These transcription and translation initiation signals may vary in "strength" as measured by the quantity of gene specific messenger RNA and protein synthesized, respectively. The DNA expression vector, which contains a promoter, may also contain any combination of various "strong" transcription and/or translation initiation signals. For instance, efficient translation in *E. coli* requires a Shine-Dalgarno ("SD" sequence about 7-9 bases 5' to the initiation codon ("ATG") to provide a ribosome binding site. Thus, any SD-ATG combination that can be utilized by host cell ribosomes may be employed. Such combinations include but are not limited to the SD-ATG combination from the *cro* gene or the N gene of coliphage lambda, or from the *E. coli* tryptophan E, D, C, B or A genes. Additionally, any SD-ATG combination produced by recombinant DNA or other techniques involving incorporation of synthetic nucleotides may be used. The present invention also relates to anti-sense nucleic acid for essential cell proteins, such as replication proteins which serve to tender host cells incapable of further cell growth and division. Anti-sense regulation has been described by Rosenberg et al., Nature, 313:703 (1985); Preiss et al., Nature, 313:27 (1985); Melton, Proc. Natl. Acad. Sci. USA, 82:144 (1985); Izaut et al., Science, 229:342 (1985); Kim et al., Cell, 42:129 (1985); Bestka et al., Proc Natl.. Acad. Sci. USA, 81:7525 (1984); Coleman et al., Cell, 37:429 (1984); and McQany et al., Proc. Natl. Acad. Sci. USA, 83:399 (1986), which are hereby incorporated by reference.

Once the isolated DNA molecules encoding the CPS1 polypeptide or iron reductase have been cloned into an expression system, they are ready to be incorporated into a host cell. Such incorporation can be carried out by the

various forms of transformation noted above, depending upon the vector host cell system. Suitable host cells include, but are not limited to, bacteria, virus, yeast, mammalian cells, insect, plant, and the like. In the present invention, the host cells are from plants such as corn, oat, grass, weeds, bamboo, and
5 sugarcane. In this aspect of the present invention, large numbers of compounds can be screened for their activity as inhibitors of CPS1 protein, iron reductase or permease/MFS transporter by a high throughput screening assay as described in U.S. Patent No. 5,767,946. Generally, a library of compounds is assayed for inhibition of an enzyme catalyzed reaction and the amounts of fluorescence
10 bound to individual suspendable solid supports measured to determine the degree of inhibition. For example, the amount of fluorescence bound to a microbead in the presence of inhibitory compounds is greater than for non-inhibitory compounds. The amounts of fluorescence bound to individual beads are determined by confocal microscopy. Using this type of assay, inhibition can be
15 determined, e.g., of a peptide synthetase such as CPS1. For CPS1 the substrate can be amino acids (or hydroxy acids), linked at one end to the microbead and at the other end to a fluorescent label. The enzyme inhibitors can be utilized to impart fungal resistance to a variety of vertebrate organisms.

Another aspect of the present invention involves using one or more of the
20 above DNA molecules encoding the CPS1 polypeptide or a gene encoding an enzyme that degrades the CPS1 product to transform organisms to impart fungal resistance to the organism. This concept of pathogen-derived resistance, according to U.S. Patent No. 5,840,481 is that host resistance to a particular parasite can effectively be engineered by introducing a gene, gene fragment, or
25 modified gene or gene fragment of the pathogen into the host. This approach is based on the fact that in any parasite-host interaction, there are certain parasite-encoded cellular functions (activities) that are essential to the parasite but not to the host and that when one of the essential functions of the parasite such as survival or reproduction is disrupted, the parasitic process will be stopped.
30 "Disruption" refers to any change that diminishes the survival, reproduction, or ineffectivity of the parasite. Such essential functions, which are under the

control of the parasite's genes, can be disrupted by the presence of a corresponding gene product in the host which is (1) dysfunctional, (2) in excess, or (3) appears in the wrong context or at the wrong developmental stage in the parasite's life cycle. If such faulty signals are designed specifically for parasitic cell functions, they will have little effect on the host. Therefore, the procedure for making organisms, for example, resistant to infection by one or more fungus involve isolating DNA coding for a gene such as CPS1 of a fungus, operably linking the DNA within an expression vector; and transforming a cell or tissue with the expression vector. The transformed cells or tissue in the presence of the fungus such as *Cochliobolus heterostrophus* where the CPS1 DNA is expressed as a gene product and the CPS protein disrupts the essential activity of the fungi.

Dosages, Formulations and Routes of Administration of the Agents of the Invention

The therapeutic agents identified by the methods of the invention may be administered at dosages of at least about 0.01 to about 100 mg/kg, more preferably about 0.1 to about 50 mg/kg, and even more preferably about 0.1 to about 30 mg/kg, of body weight, although other dosages may provide beneficial results. The amount administered will vary depending on various factors including, but not limited to, the agent chosen, the disease, whether prevention or treatment is to be achieved, and if the agent is modified for bioavailability and *in vivo* stability.

Administration of a sense or antisense nucleic acid molecule encoding a therapeutic agent may be accomplished through the introduction of cells transformed with an expression cassette comprising the nucleic acid molecule (see, for example, WO 93/02556) or the administration of the nucleic acid molecule (see, for example, Felgner et al., U.S. Patent No. 5,580,859, Pardoll et al., *Immunity*, 3:165 (1995); Stevenson et al., *Immunol. Rev.*, 145:211 (1995); Molling, *J. Mol. Med.*, 75:242 (1997); Donnelly et al., *Ann. N.Y. Acad. Sci.*, 772:40 (1995); Yang et al., *Mol. Med. Today*, 2:476 (1996); Abdallah et al., *Biol. Cell*, 85:1 (1995)). Pharmaceutical formulations, dosages and routes of

administration for nucleic acids are generally disclosed, for example, in Felgner et al., *supra*.

The therapeutic agents of the invention are amenable to chronic use for prophylactic purposes, preferably by systemic administration.

5 Administration of the therapeutic agents in accordance with the present invention may be continuous or intermittent, depending, for example, upon the recipient's physiological condition, whether the purpose of the administration is therapeutic or prophylactic, and other factors known to skilled practitioners. The administration of the agents of the invention may be essentially continuous over
10 a preselected period of time or may be in a series of spaced doses. Both local and systemic administration is contemplated.

One or more suitable unit dosage forms comprising the therapeutic agents of the invention, which, as discussed below, may optionally be formulated for sustained release, can be administered by a variety of routes including oral, or
15 parenteral, including by rectal, buccal, vaginal and sublingual, transdermal, subcutaneous, intravenous, intramuscular, intraperitoneal, intrathoracic, intrapulmonary and intranasal routes. The formulations may, where appropriate, be conveniently presented in discrete unit dosage forms and may be prepared by any of the methods well known to pharmacy. Such methods may include the
20 step of bringing into association the therapeutic agent with liquid carriers, solid matrices, semi-solid carriers, finely divided solid carriers or combinations thereof, and then, if necessary, introducing or shaping the product into the desired delivery system.

When the therapeutic agents of the invention are prepared for oral
25 administration, they are preferably combined with a pharmaceutically acceptable carrier, diluent or excipient to form a pharmaceutical formulation, or unit dosage form. The total active ingredients in such formulations comprise from 0.1 to 99.9% by weight of the formulation. By "pharmaceutically acceptable" it is meant the carrier, diluent, excipient, and/or salt must be compatible with the
30 other ingredients of the formulation, and not deleterious to the recipient thereof. The active ingredient for oral administration may be present as a powder or as

granules; as a solution, a suspension or an emulsion; or in achievable base such as a synthetic resin for ingestion of the active ingredients from a chewing gum. The active ingredient may also be presented as a bolus, electuary or paste.

Formulations suitable for vaginal administration may be presented as
5 pessaries, tampons, creams, gels, pastes, douches, lubricants, foams or sprays containing, in addition to the active ingredient, such carriers as are known in the art to be appropriate. Formulations suitable for rectal administration may be presented as suppositories.

Pharmaceutical formulations containing the therapeutic agents of the
10 invention can be prepared by procedures known in the art using well-known and readily available ingredients. For example, the agent can be formulated with common excipients, diluents, or carriers, and formed into tablets, capsules, suspensions, powders, and the like. Examples of excipients, diluents, and carriers that are suitable for such formulations include the following fillers and
15 extenders such as starch, sugars, mannitol, and silicic derivatives; binding agents such as carboxymethyl cellulose, HPMC and other cellulose derivatives, alginates, gelatin, and polyvinyl-pyrrolidone; moisturizing agents such as glycerol; disintegrating agents such as calcium carbonate and sodium bicarbonate; agents for retarding dissolution such as paraffin; resorption
20 accelerators such as quaternary ammonium compounds; surface active agents such as cetyl alcohol, glycerol monostearate; adsorptive carriers such as kaolin and bentonite; and lubricants such as talc, calcium and magnesium stearate, and solid polyethyl glycols.

For example, tablets or caplets containing the agents of the invention can
25 include buffering agents such as calcium carbonate, magnesium oxide and magnesium carbonate. Caplets and tablets can also include inactive ingredients such as cellulose, pregelatinized starch, silicon dioxide, hydroxy propyl methyl cellulose, magnesium stearate, microcrystalline cellulose, starch, talc, titanium dioxide, benzoic acid, citric acid, corn starch, mineral oil, polypropylene glycol,
30 sodium phosphate, and zinc stearate, and the like. Hard or soft gelatin capsules containing an agent of the invention can contain inactive ingredients such as

gelatin, microcrystalline cellulose, sodium lauryl sulfate, starch, talc, and titanium dioxide, and the like, as well as liquid vehicles such as polyethylene glycols (PEGs) and vegetable oil. Moreover, enteric coated caplets or tablets of an agent of the invention are designed to resist disintegration in the stomach and
5 dissolve in the more neutral to alkaline environment of the duodenum.

The therapeutic agents of the invention can also be formulated as elixirs or solutions for convenient oral administration or as solutions appropriate for parenteral administration, for instance by intramuscular, subcutaneous or intravenous routes.

10 The pharmaceutical formulations of the therapeutic agents of the invention can also take the form of an aqueous or anhydrous solution or dispersion, or alternatively the form of an emulsion or suspension.

Thus, the therapeutic agent may be formulated for parenteral administration (e.g., by injection, for example, bolus injection or continuous
15 infusion) and may be presented in unit dose form in ampules, pre-filled syringes, small volume infusion containers or in multi-dose containers with an added preservative. The active ingredients may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively,
20 the active ingredients may be in powder form, obtained by aseptic isolation of sterile solid or by lyophilization from solution, for constitution with a suitable vehicle, e.g., sterile, pyrogen-free water, before use.

These formulations can contain pharmaceutically acceptable vehicles and adjuvants which are well known in the prior art. It is possible, for example, to
25 prepare solutions using one or more organic solvent(s) that is/are acceptable from the physiological standpoint, chosen, in addition to water, from solvents such as acetone, ethanol, isopropyl alcohol, glycol ethers such as the products sold under the name "Dowanol", polyglycols and polyethylene glycols, C₁-C₄ alkyl esters of short-chain acids, preferably ethyl or isopropyl lactate, fatty acid
30 triglycerides such as the products marketed under the name "Miglyol", isopropyl myristate, animal, mineral and vegetable oils and polysiloxanes.

The compositions according to the invention can also contain thickening agents such as cellulose and/or cellulose derivatives. They can also contain gums such as xanthan, guar or carbo gum or gum arabic, or alternatively polyethylene glycols, bentones and montmorillonites, and the like.

5 It is possible to add, if necessary, an adjuvant chosen from antioxidants, surfactants, other preservatives, film-forming, keratolytic or comedolytic agents, perfumes and colorings. Also, other active ingredients may be added, whether for the conditions described or some other condition.

 For example, among antioxidants, t-butylhydroquinone, butylated
10 hydroxyanisole, butylated hydroxytoluene and α -tocopherol and its derivatives may be mentioned. The galenical forms chiefly conditioned for topical application take the form of creams, milks, gels, dispersion or microemulsions, lotions thickened to a greater or lesser extent, impregnated pads, ointments or sticks, or alternatively the form of aerosol formulations in spray or foam form or
15 alternatively in the form of a cake of soap.

 Additionally, the agents are well suited to formulation as sustained release dosage forms and the like. The formulations can be so constituted that they release the active ingredient only or preferably in a particular part of the intestinal or respiratory tract, possibly over a period of time. The coatings,
20 envelopes, and protective matrices may be made, for example, from polymeric substances, such as polylactide-glycolates, liposomes, microemulsions, microparticles, nanoparticles, or waxes. These coatings, envelopes, and protective matrices are useful to coat indwelling devices, e.g., stents, catheters, peritoneal dialysis tubing, and the like.

25 The therapeutic agents of the invention can be delivered via patches for transdermal administration. See U.S. Patent No. 5,560,922 for examples of patches suitable for transdermal delivery of a therapeutic agent. Patches for transdermal delivery can comprise a backing layer and a polymer matrix which has dispersed or dissolved therein a therapeutic agent, along with one or more
30 skin permeation enhancers. The backing layer can be made of any suitable material which is impermeable to the therapeutic agent. The backing layer

serves as a protective cover for the matrix layer and provides also a support function. The backing can be formed so that it is essentially the same size layer as the polymer matrix or it can be of larger dimension so that it can extend beyond the side of the polymer matrix or overlay the side or sides of the polymer matrix and then can extend outwardly in a manner that the surface of the extension of the backing layer can be the base for an adhesive means. Alternatively, the polymer matrix can contain, or be formulated of, an adhesive polymer, such as polyacrylate or acrylate/vinyl acetate copolymer. For long-term applications it might be desirable to use microporous and/or breathable backing laminates, so hydration or maceration of the skin can be minimized.

Examples of materials suitable for making the backing layer are films of high and low density polyethylene, polypropylene, polyurethane, polyvinylchloride, polyesters such as poly(ethylene phthalate), metal foils, metal foil laminates of such suitable polymer films, and the like. Preferably, the materials used for the backing layer are laminates of such polymer films with a metal foil such as aluminum foil. In such laminates, a polymer film of the laminate will usually be in contact with the adhesive polymer matrix.

The backing layer can be any appropriate thickness which will provide the desired protective and support functions. A suitable thickness will be from about 10 to about 200 microns.

Generally, those polymers used to form the biologically acceptable adhesive polymer layer are those capable of forming shaped bodies, thin walls or coatings through which therapeutic agents can pass at a controlled rate. Suitable polymers are biologically and pharmaceutically compatible, nonallergenic and insoluble in and compatible with body fluids or tissues with which the device is contacted. The use of soluble polymers is to be avoided since dissolution or erosion of the matrix by skin moisture would affect the release rate of the therapeutic agents as well as the capability of the dosage unit to remain in place for convenience of removal.

Exemplary materials for fabricating the adhesive polymer layer include polyethylene, polypropylene, polyurethane, ethylene/propylene copolymers,

ethylene/ethylacrylate copolymers, ethylene/vinyl acetate copolymers, silicone elastomers, especially the medical-grade polydimethylsiloxanes, neoprene rubber, polyisobutylene, polyacrylates, chlorinated polyethylene, polyvinyl chloride, vinyl chloride-vinyl acetate copolymer, crosslinked polymethacrylate polymers (hydrogel), polyvinylidene chloride, poly(ethylene terephthalate), butyl rubber, epichlorohydrin rubbers, ethylenvinyl alcohol copolymers, ethylene-vinyloxyethanol copolymers; silicone copolymers, for example, polysiloxane-polycarbonate copolymers, polysiloxanepolyethylene oxide copolymers, polysiloxane-polymethacrylate copolymers, polysiloxane-alkylene copolymers (e.g., polysiloxane-ethylene copolymers), polysiloxane-alkylenesilane copolymers (e.g., polysiloxane-ethylenesilane copolymers), and the like; cellulose polymers, for example methyl or ethyl cellulose, hydroxy propyl methyl cellulose, and cellulose esters; polycarbonates; polytetrafluoroethylene; and the like.

Preferably, a biologically acceptable adhesive polymer matrix should be selected from polymers with glass transition temperatures below room temperature. The polymer may, but need not necessarily, have a degree of crystallinity at room temperature. Cross-linking monomeric units or sites can be incorporated into such polymers. For example, cross-linking monomers can be incorporated into polyacrylate polymers, which provide sites for cross-linking the matrix after dispersing the therapeutic agent into the polymer. Known cross-linking monomers for polyacrylate polymers include polymethacrylic esters of polyols such as butylene diacrylate and dimethacrylate, trimethylol propane trimethacrylate and the like. Other monomers which provide such sites include allyl acrylate, allyl methacrylate, diallyl maleate and the like.

Preferably, a plasticizer and/or humectant is dispersed within the adhesive polymer matrix. Water-soluble polyols are generally suitable for this purpose. Incorporation of a humectant in the formulation allows the dosage unit to absorb moisture on the surface of skin which in turn helps to reduce skin irritation and to prevent the adhesive polymer layer of the delivery system from failing.

Therapeutic agents released from a transdermal delivery system must be capable of penetrating each layer of skin. In order to increase the rate of permeation of a therapeutic agent, a transdermal drug delivery system must be able in particular to increase the permeability of the outermost layer of skin, the stratum corneum, which provides the most resistance to the penetration of molecules. The fabrication of patches for transdermal delivery of therapeutic agents is well known to the art.

For administration to the upper (nasal) or lower respiratory tract by inhalation, the therapeutic agents of the invention are conveniently delivered from an insufflator, nebulizer or a pressurized pack or other convenient means of delivering an aerosol spray. Pressurized packs may comprise a suitable propellant such as dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol, the dosage unit may be determined by providing a valve to deliver a metered amount.

Alternatively, for administration by inhalation or insufflation, the composition may take the form of a dry powder, for example, a powder mix of the therapeutic agent and a suitable powder base such as lactose or starch. The powder composition may be presented in unit dosage form in, for example, capsules or cartridges, or, e.g., gelatine or blister packs from which the powder may be administered with the aid of an inhalator, insufflator or a metered-dose inhaler.

For intra-nasal administration, the therapeutic agent may be administered via nose drops, a liquid spray, such as via a plastic bottle atomizer or metered-dose inhaler. Typical of atomizers are the Mistometer (Wintrop) and the Medihaler (Riker).

The local delivery of the therapeutic agents of the invention can also be by a variety of techniques which administer the agent at or near the site of disease. Examples of site-specific or targeted local delivery techniques are not intended to be limiting but to be illustrative of the techniques available. Examples include local delivery catheters, such as an infusion or indwelling

catheter, e.g., a needle infusion catheter, shunts and stents or other implantable devices, site specific carriers, direct injection, or direct applications.

For topical administration, the therapeutic agents may be formulated as is known in the art for direct application to a target area. Conventional forms for this purpose include wound dressings, coated bandages or other polymer coverings, ointments, creams, lotions, pastes, jellies, sprays, and aerosols, as well as in toothpaste and mouthwash, or by other suitable forms, e.g., via a coated condom. Ointments and creams may, for example, be formulated with an aqueous or oily base with the addition of suitable thickening and/or gelling agents. Lotions may be formulated with an aqueous or oily base and will in general also contain one or more emulsifying agents, stabilizing agents, dispersing agents, suspending agents, thickening agents, or coloring agents. The active ingredients can also be delivered via iontophoresis, e.g., as disclosed in U.S. Patent Nos. 4,140,122; 4,383,529; or 4,051,842. The percent by weight of a therapeutic agent of the invention present in a topical formulation will depend on various factors, but generally will be from 0.01% to 95% of the total weight of the formulation, and typically 0.1-25% by weight.

When desired, the above-described formulations can be adapted to give sustained release of the active ingredient employed, e.g., by combination with certain hydrophilic polymer matrices, e.g., comprising natural gels, synthetic polymer gels or mixtures thereof.

Drops, such as eye drops or nose drops, may be formulated with an aqueous or non-aqueous base also comprising one or more dispersing agents, solubilizing agents or suspending agents. Liquid sprays are conveniently delivered from pressurized packs. Drops can be delivered via a simple eye dropper-capped bottle, or via a plastic bottle adapted to deliver liquid contents dropwise, via a specially shaped closure.

The therapeutic agent may further be formulated for topical administration in the mouth or throat. For example, the active ingredients may be formulated as a lozenge further comprising a flavored base, usually sucrose and acacia or tragacanth; pastilles comprising the composition in an inert base

such as gelatin and glycerin or sucrose and acacia; mouthwashes comprising the composition of the present invention in a suitable liquid carrier; and pastes and gels, e.g., toothpastes or gels, comprising the composition of the invention.

The formulations and compositions described herein may also contain
5 other ingredients such as antimicrobial agents, or preservatives. Furthermore, the active ingredients may also be used in combination with other therapeutic agents, for example, oral contraceptives, bronchodilators, anti-viral agents, steroids and the like.

The invention will be further described by the following non-limiting
10 examples.

15

Example 1

Mutant Preparation and Characterization

Materials and Methods

Strains, Media, Crosses and Transformation. C4 (*ToxI*⁺; MAT-2) and C5 (*ToxI*⁻; MAT-1) are members of near-isogenic *C. heterostrophus* strains (Leach et al., 1982, *supra*). R.C4.2696 (*Tox*⁺; MAT-2; *hygB*^R) is a C4-derived mutant generated using the REMI mutagenesis procedure (Lu et al., Proc. Natl. Acad. Sci. USA, 91:12649 (1994)). Strains 1301R33 (*Tox*⁻; MAT-2; *hygB*^R), 1301R45 (*Tox*⁻; MAT-1; *hygB*^R) 1301R26 (*Tox*⁺; MAT-2; *hygB*^R) are progeny of the cross C5 X R.C4.2696. Culture media, including CM (complete medium), CMX (complete medium with xylose instead of glucose), CMNS (CM with salts omitted), and MM (minimal medium) have been described, as have mating procedures (Leach et al., 1982, *supra*; Turgeon et al., Mol. Gen. Genet., 201:450 (1985)). All strains were grown at 24°C under the warm white light or black light (F40/350BL) (Sylvania Inc., Danvers, MA). Ascospore germination was
20 done at 32°C in the dark for 3 days. REMI transformants were purified by transferring the transformants from the original REMI plates to fresh CMNS
25
30

medium containing hygromycin B (Calbiochem^R) at 80 µg/ml. For conidiation, stable transformants were transferred to CMX containing the same drug but at a higher concentration (120 µg/ml) to compensate for reduced drug activity due to the inhibition by the salts in the medium. Single conidia were picked up under a dissecting microscope and grown on CMNS hygromycin B plates; stable colonies were then transferred to individual CMX/hygromycin plates. All purified transformants were stored at -70°C in CM liquid medium containing 25% of glycerol in 96-well microtiter dishes.

Bioassays. Fungal strains were grown on CMX plates (100 X 15 mm) for 7-10 days at 24°C under the light for maximum conidiation. To verify normal T-toxin production by a race T isolate, 1.0 ml of T-toxin-sensitive *E. coli* (DH5a) cells were evenly spread on LB medium containing ampicillin (100 µg/ml) and the plates were allowed to air dry for 30 minutes in a laminar hood. Agar plugs bearing fungal mycelia were inoculated (upside down) onto the *E. coli* cell lawn and the plates were incubated at 32°C. Wild type race T and race O were used as controls for each assay plate. T-toxin-producing strains of the fungus will inhibit growth of the *E. coli* cells and produce halos. *Tox*⁻ mutants can be distinguished from wild type by failure to produce a halo (tight) or by production of halos smaller (leaky) or larger than wild type (overproducing). All *Tox*⁻ mutants were transferred to Fries medium (Pringle et al., Phytopathology, 47:369 (1957)), which optimizes toxin production, and retested.

T-cytoplasm corn plants (inbred W64A) are used to verify the *Tox*⁻ mutants identified from the *E. coli* assay using the procedure described below. Mutants defective in T-toxin production fail to produce typical race T symptoms on T-corn. Pathogenicity phenotype on N-cytoplasm corn and virulence of *Tox*⁺ strains to T-cytoplasm corn were determined by a plant assay where, about 3,000 transformants generated using the REMI mutagenesis procedure (Lu et al., Proc. Natl. Acad. Sci. USA, 91:12649 (1994)) were screened for mutants defective in ability to cause disease on corn plants. Two week old N-cytoplasm corn plants (inbred W64A) grown in the green house (5-6 plants in one 4" X 6" pot) were inoculated with 5 ml conidial suspensions (10⁵ conidia/ml) using a pressurized

Preval Spray Gun Power Unit thin layer chromatography sprayer (Alltech Associates, Deerfield, IL), incubated in the mist chamber for 24 hours (23 °C) and then taken to the growth chamber (23 °C, 80% humidity, 14 hours of light). The mutant phenotypes were determined by occurrence of apparent variations in disease symptom development, mainly by lesion size comparison. Mutants producing lesions smaller than wild type were retested and lengths of typical lesions from each mutant were compared with wild type 7 days after inoculation and measurements were taken for statistical evaluation.

DNA manipulations and sequencing. Genomic and plasmid DNA preparation, restriction enzyme digestions, gel electrophoresis and gel blot analysis were done using standard protocols (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (1989)). DNA was sequenced at the Cornell DNA Sequencing Facility using TaqCycle automated sequencing with DyeDeoxy terminators (Applied Biosystems, Foster City, CA). pUCATPH was used for subcloning (Table 1). Primers used for sequencing (Table 2) were designed using Primer Select (DNASTAR Inc., LaserGene System) and synthesized by the Cornell Oligonucleotide Synthesis Facility. Sequencing of each plasmid clone was initiated with vector-specific primers or primers designed to previously determined sequences. Sequences obtained were analyzed using the same system and nucleotide or protein database searches were performed with the BLAST program (Altschul et al., J. Mol. Biol., 215:403 (1990)).

25

Table 1. Transformation vectors and clones used.

Plasmid	Length (kb) ^a	Characteristics (See U.S. application Serial Nos. 60/252,649 and 60/252,732)
pUCATPH	5.1	See Figure 14 in U.S. application Serial No. 60/252,649.
PUCATPHN	4.6	Cloning vector, same as pUCATPH but lacking a 420 bp <i>NarI</i> fragment containing the <i>HindIII</i> site
p214B7	<u>9.2</u>	A clone containing pUCATPH recovered from the tagged site in mutant R.C4.2696 by religation of <i>BglIII</i> -digested genomic DNA
p214M1	<u>6.3</u>	As above but with <i>MscI</i> -digested genomic DNA
p214S1	<u>9.3</u>	As above but with <i>SacI</i> -digested genomic DNA
p214S1N	<u>3.3</u>	<i>NarI</i> fragment derived from 214S1 containing a 0.8 kb <i>NarI</i> - <i>SacI</i> fragment of genomic DNA ligated to pUC18
p214SNP	<u>8.4</u>	Vector for targeted integration constructed by ligating <i>HindIII</i> -digested pUCATPH into the <i>HindIII</i> site of p214S1N
p118BSP	<u>7.3</u>	Vector for targeted integration constructed by ligation of a 2.2 kb <i>SacI</i> fragment of p118BC4 into the <i>SacI</i> site of pUCATPH
p118BCS	<u>5.4</u>	Vector for targeted integration constructed by ligation of a 0.8 kb <i>SspI</i> fragment of p118BC4 into the <i>SspI</i> site of pUCATPHN
p118B14	<u>10.4</u>	A clone recovered from the p214SNP integration site in transformant #f118 by ligation of a <i>BglIII</i> -digested genomic DNA fragment containing the entire vector
p118BC4	<u>6.7</u>	A clone recovered from same site as above but by ligation of a <i>BclI</i> -digested genomic DNA fragment containing part of vector (214SNP) sequence

Plasmid	Length (kb) ^a	Characteristics (See U.S. application Serial Nos. 60/252,649 and 60/252,732)
p9P2	<u>7.3</u>	A clone recovered from the p118BSP integration site in transformant #9 by ligation of a <i>Pst</i> I-digested genomic DNA fragment containing pUC18
p12H6	<u>8.0</u>	A clone recovered from the p118BCS integration site in transformant #12 by ligation of a <i>Hind</i> III-digested genomic DNA fragment containing the entire vector.

^a An underlined kb number indicates that the plasmid carries genomic DNA sequences.

- 5 Table 2. Primers used for sequencing recovered genomic DNA flanking the REMI insertion site at the R.C4 2696 mutation.

Name ^a	Position ^b	Sequence ^c	Plasmid ^d	Origin ^e
M13RMT		SEQ ID NO:4	A	pUC18
1. RP1b	775	SEQ ID NO:5	A	214B7TrpC
2. RP2	604	SEQ ID NO:6	A	214B7RP1b
3. RP3	119	SEQ ID NO:7	A	214B7RP2
4. RP4	-232	SEQ ID NO:8	A	214B7RP3
5. RP5	-812	SEQ ID NO:9	A	214B7RP4
6. RP5b	-1215	SEQ ID NO:10	A	214B7RP4
7. RP6	-1392	SEQ ID NO:11	A	214B7RP5
8. RP7	-1839	SEQ ID NO:12	A	214B7RP6
TrpC		SEQ ID NO:13	A	PUCATPH
9. FP1	1885	SEQ ID NO:14	A	214B7TrpC
10. FP1b	1828	SEQ ID NO:15	B	214B7TrpC
11. FP2	2028	SEQ ID NO:16	B	214M1FP1b
12. FP3	2490	SEQ ID NO:17	C	214M1FP2
13. FP4	2949	SEQ ID NO:18	C	214S1FP3
14. FP4B	2745	SEQ ID NO:19	C	214S1FP4
15. FP5	3421	SEQ ID NO:20	C	214S1FP4
16. FP6	3948	SEQ ID NO:21	C	214S1FP5
17. FP7	4411	SEQ ID NO:22	C, D	214S1FP6
18. FP8	5035	SEQ ID NO:23	D	118B14FP7

Name ^a	Position ^b	Sequence ^c	Plasmid ^d	Origin ^e
19. FP9	5457	SEQ ID NO:24		118BC4FP8
20. RP48	2865	SEQ ID NO:25	D	214S1FP6
21. FP10	5790	SEQ ID NO:26	F	9P2FP9
22. FP11	6327	SEQ ID NO:27	F	9P2FP10
23. FP11b	6211	SEQ ID NO:28	F	9P2FP10
24. FP12	6457	SEQ ID NO:29	F	9P2FP11
25. FP13	6854	SEQ ID NO:30	F	9P2FP12
26. FP14	7400	SEQ ID NO:31	F	9P2FP13
27. FP15	7771	SEQ ID NO:32	F	9P2FP14
28. FP16	8145	SEQ ID NO:33	F	9P2FP15
29. FP17	8492	SEQ ID NO:34	F	9P2FP16
M13F40		SEQ ID NO:35	G	pUC18
30. RP1	8953	SEQ ID NO:36	G	9P5M13F4
31. RP2	8559	SEQ ID NO:37	G	9P5RP1

- ^a “RP” indicates reverse primer; “FP” indicates forward primer. Primers designed to genomic DNA sequences are numbered in order. Primers 1-17 have a leading number “214”; 18-20 with “118”; 21-29 with “9P2” and 30-31 with “9P5”. M13RMT (a M13R mutant version; there is a mutation in the polylinker of pUC18) and M13F-40 were provided by Cornell DNA Sequencing Facility. *TrpC* primer site is in the pUCATPH *TrpC* promoter region 38 bp from *SaII* site with sequencing direction from *SaII* to *KpnI*.
- ^b The position of the first base of each primer corresponds to the assembled sequence (*CPSI* + *TES1*, total 11.3 kb).
- ^c Each primer sequence is given in the 5' to 3' direction.
- ^d Plasmids used as templates for each sequencing reaction. A= p214B7; B=p214M1; C=p214S1; D=p118B14; E=p118BC4; F=p9P2; G=p9P5 (=9P2)
- ^e Original sequences that were used for primer design.

Results

Recovery of tagged DNA from the REMI insertion site and targeted gene disruption. Genomic DNA of mutant R.C4.2696 was digested with *BglIII*, *MscI* (no sites in pUCATPH) or *SacI* (which cuts the vector once) and purified by phenol extraction and ethanol precipitation, then dissolved in TE (pH 8.0). Ligation was performed in 50 µl reaction mixture, containing 1 x T4 DNA ligase buffer with 10 mM ATP, 60 units T4 DNA ligase (New England Biolabs, Beverly, MA) and 3 µg of *BglIII*-digested genomic DNA, at 14°C overnight. Ten µl of ligation mixture was used to transform 200 µl of competent DH5α cells,

prepared using the calcium chloride treatment (Sambrook et al., 1989, *supra*) to ampicillin resistance. Ampicillin resistant clones were analyzed by digestion of plasmid DNA with several diagnostic restriction enzymes and clones containing the REMI vector plus flanking genomic DNA were sequenced using the vector-specific primers (M13R or *TrpC*). Three plasmids, p214B7, p214M1 and p214S1 were recovered and used for sequencing. p214B7 contains 4.2 kb flanking DNA (3.4 left; 0.7 right); p214M1 contains 0.1 kb left flank that overlaps with p214B7 and 1.1 kb right flank that overlaps with p214S1, which contains 3.2 kb flanking DNA on the left only.

For targeted gene disruption in wild type, p214B7 was amplified and plasmid DNA purified by equilibrium centrifugation in CsCl-ethidium bromide gradients (Sambrook et al., 1989, *supra*). Thirty µg of plasmid DNA (linearized with *Bgl*III for double crossover integration) were used to transform wild type and the transformants were purified by isolation of single conidia, assayed for pathogenicity and characterized by gel blot analysis.

Sequence extension by targeted integration and plasmid rescue. Two overlapping cosmid clones were isolated by probing a genomic DNA library of C4 constructed on a cosmid vector, but both extended into the left region only of p214B7. To extend to the right, a chromosome walking strategy was employed. Three targeted gene disruption experiments (each followed by plasmid rescue) were done successively. In the first experiment, a vector was constructed as follows: p214S1 was digested with *Nar*I and religated to create p214S1N, which was then digested with *Hind*III and ligated into the *Hind*III site of pUCATPH to create p214SNP for transformation of race O (C5). One transformant (Tx118) resulting from homologous integration (confirmed by gel blot analysis) was used for plasmid rescue as described above. Two new plasmids p118B14 and p118BC4 were recovered, both of which carry sequence at the 3' end but only 172 and 680 bp more than p214S1, respectively. To continue the walk, p118B14 was digested with *Sac*I and ligated into the *Sac*I site of pUCATPH to create p118BSP. This vector was linearized with *Bgl*III and transformed into wild type and one plasmid, p9P2 was recovered (from transformant Tx9), which extends

4.4 kb into the region 3' of p118BC4 and contains the 3' end of *CPS1*. The recovered plasmid p9P2 includes the entire pUC18 sequence on p118BSP and 4.6 kb of genomic DNA that contains all of ORF1 (*CPS1*), including the stop codon (TAG) and 3.0 kb of genomic region 3' of the stop codon. A third
5 experiment was done in an attempt to recover a 15 kb *XhoI* fragment at the 3' end of that tagged gene. p118BCS was constructed by subcloning a 0.8 kb *SspI* fragment into the same site pUCATPHN. Plasmid rescue using *XhoI* digested-genomic DNA of a transformant (TXI2) failed to recover the 15 kb *XhoI* fragment, but p12H6 was recovered using *HindIII*-digested genomic DNA of the
10 same transformant; the genomic DNA matched that already cloned on p9P2.

Characterization of the REMI mutant. In all culture conditions used, mutant R.C4.2696 grew just like wild type with no variations in growth rate, color and morphological features. It produces normal appressorium-forming conidia that germinate and form infection structures like wild type when induced
15 on artificial surfaces and shows normal mating ability when crossed to wild type testers. No pleiotropic phenotypes associated with the mutation have been detected so far. The mutant differs from wild type in the ability to cause disease on corn plants.

The lengths of 100 typical lesions from corn leaves inoculated with wild
20 type race O and a mutant progeny R45 (*Tox*⁻, *hygB*^R) carrying the R.C4.2696 mutation were measured 7 days after inoculation and values plotted.

When tested on T-cytoplasm corn, the mutant produces race T type symptoms but the disease develops more slowly than with wild type although it produces wild type levels of T-toxin as detected in a microbial assay, suggesting that the
25 reduced virulence is not related to a deficiency in the ability to produce T-toxin. This is clearer on N-cytoplasm corn where the mutant produces lesions significantly smaller than those produced by wild type. When the mutant was crossed to a wild type race O tester, the small lesion phenotype and ability to produce T-toxin segregated independently, indicating that mutant phenotype is
30 not associated with the reduced fitness trait tightly linked with the *Tox1* locus (Klittich et al., Phytopathology, 76:1294 (1986)). The statistical evaluation of

lesion size in the wild type race O genetic background indicates that the mutation causes 60% reduction in the fungal virulence to corn plants. Table 3 depicts the statistical analysis that 86% of the mutant lesions are less than 4 mm in length (average size of 3.5 mm), 60% reduced compared to that of wild type (8.5 mm).

5

Table 3

<u>Strain</u>	<u>Frequency</u>			<u>Lesion size (mm)</u>		
	1-4	5-8	9-12	<u>Mean</u>	<u>SD</u>	
WT	0	52	48	8.5	1.0	A*
R45	86	14	0	3.5	0.9	B

*Significant difference at $P < 0.01$.

10 The mutant phenotype is caused by a tagged, single site mutation. In crosses between the mutant and wild type testers, progeny segregated 1:1 for parental types only and all hygromycin B-resistant progeny produced lesions similar to the mutant parent; all hygromycin B-sensitive progeny produced wild type lesions, indicating that a tagged mutation is responsible for the reduced pathogenicity of the mutant. Table 4 depicts the progeny segregation data.

15

Table 4

Cross	Progeny	path	Parental type		Nonparental type	
			PATH hygB ^R	path hygB ^S	PATH hygB ^R	path hygB ^S
R.C4.2696 x C5	random spores		24	22	0	0
1301-R33* x C5	tetrad1		4	4	0	0
	tetrad2		4	4	0	0
	tetrad3		4	4	0	0
	Random spores		21	22	0	0

35

*13012-R33 (path, hygB^R, Tox⁻, MAT-2) is a progeny from the first cross, carrying the R.C4.2696 mutation.

Example 2

Cloning, Sequencing and Characterization of DNA Flanking the REMI Vector Insertion Site

5 A total of 11.3 kb of genomic DNA surrounding the insertion site was
cloned and completely sequenced (SEQ ID NO:59; Figure 2). The sequence was
derived from seven plasmid clones. The first three (p2l4B7, p2l4Ml and p2l4Sl)
were recovered from the tagged site in mutant R.C4.2696 and cover about 60%
(6.6 kb) of the entire region. The rest (p118B14, p118BC4, p9P2 and p12H6)
10 were recovered from transformants generated using the chromosome walking
strategy. DNA to the left of the insertion site (3.4 kb) was cloned on p2l4B7;
DNA on the right (7.9 kb) was cloned on different overlapping plasmids. p9P2
carries the largest amount (4.6 kb) including genomic DNA on p12H6.

 Analysis of the combined sequences revealed two open reading frames
15 (ORFs). ORF1 (5.4 kb) starts 576 bp upstream of the REMI vector insertion site
and ends with an in-frame stop codon (TAG) 3029 bp from the end of the
sequenced region in the right flank. No "TATA" box-like element is found in
the expected position, but five putative "CAAT" boxes are located upstream of
the start codon (ATG), three of them are in the range found in most filamentous
20 fungal promoters (60-200 bp) (Gurr et al., 1987, *infra*). Sequence around ATG
of ORF1 (CACCATGCT) (SEQ ID NO:38) is similar to the fungal consensus
(CACCATGGC) (SEQ ID NO:39). Although there are several ATGs found
upstream, they are less likely to be used as a start codon because the surrounding
sequences lack similarity to the consensus. Three putative introns are identified
25 by their conserved 5' and 3' border sequences and potential branch sites (Table
5). Splicing these introns eliminated stop codons which would otherwise
interrupt the 5.4 kb open reading frame. Three introns have similar size (45-53
bp respectively) which is in the range of intron size determined from most fungal
genes. A putative polyadenylation signal (ATAA) is found 223 bp downstream
30 of the translation termination site.

The G+C content of ORF1 is 51.5%, which is similar to most

Cochliobolus genes (Turgeon et al., Mol. Gen. Gene., 238:270 (1993); VanWert et al., Curr. Genet., 22:29 (1992); Yang et al., Plant Cell, 8:2139 (1996); Rose et al., 1996, *supra*). Interestingly, ORF1 is flanked by two regions of G+C rich DNA. The first (1.4 kb, 60.3% G+C) is found between ORF1 and ORF2; the
5 second (1.2 kb, 60.3% G+C) is found 1.8 kb downstream of the stop codon of ORF1. Database searches using the translated protein sequence of ORF1 revealed high similarity to SafB, one of the multifunctional enzymes catalyzing the biosynthesis of the cyclic peptide antibiotic saframycin Mx1 produced by the bacterium *Myxococcus xanthus* (Pospiech et al., Microbiology, 142:741 (1996)).

10 The entire nucleotide sequence of ORF1 (CPS1) is designated SEQ ID NO:2 (6,550 base pairs from the 11.3 kb sequenced region, Figure 2). The deduced amino acid sequence of CPS1 protein is designated SEQ ID NO:3. A modification of the ChCPS1 sequence, including changes in three base pairs ("ATG" added between positions 5349 and 5350 of the GenBank entry
15 (GenBank Accession number AF332878)) and an addition of 31 amino acids (the first thirty amino acids ("MMGNYAFNPDNQSYDGQFGSPGEASRRST") were added at the N-terminus based on the selection of a new start codon and an additional methionine ("M" at position 1489 was missing in the Genbank entry)) is designated SEQ ID NO:50 (6553 base pairs). The deduced amino acid
20 sequence of the modified ChCPS1 protein is designated SEQ ID NO:185 (1774 amino acids; revised version of the original CPS1 protein (GenBank Accession number AAG53991)). The open reading frame is 5,474 base pairs (736-6209), a 93 base pair increase compared to the deposited sequence that was 5,381 bp. A new start codon (position 736, the original one at position 826) was proposed
25 based on the amino acid alignment of several CPS1 orthologs from different fungi that revealed conserved residues in this region. The stop codon (6,209) is the same as the original GenBank sequence.

Table 5. Characteristics of putative introns in *CPS1* and *TES1*

Gene	Intron	Size (bp)	Location	5' Border	3' Border	Branch Site
CPS1	I	45	3060-3105	GTAAGT	TAG	GTCTAAC
	II	51	4532-4582	GTAAGT	CAG	TGCTAAC
	III	53	5187-5239	GTACGT	CAG	TACTAAC
TES1	I	49	528-566	GTAAGT	TAG	CCTTAAG
Cons				GTA ^A /C ^{GT}	T/C ^{AG}	YNCTAAC*

ORF2 starts about 1.6 kb upstream of the start codon of *CPS1* and is transcribed in the opposite direction (Figure 2). No "TATA" box-like element and CAAT box are found; instead, an AT-rich sequence "AAACTAT" is located 11 bp upstream of the start codon ATG and a CT motif is found in the -30 region, which is characteristic of a number of fungal genes that lack a CAAT box in their promoter region (Gurr et al., In: Gene Structure in Eukaryotic Microbes, Vol.22, published by the Society for General Microbiology, Oxford, England: IRL Press, Kinghorn, ed., pp 93-140 (1987)). The sequence around ATG matches perfectly fungal gene consensus. A putative intron (50 bp) is found in the middle of ORF2 with conserved 5' and 3' border sequences and a potential branch site (Table 5). A putative polyadenylation signal (AAATA) is found 189 bp downstream of the translation stop codon TGA. The G+C content of ORF2 is 55.5%, which is slightly higher than the normal range because the 5' end of ORF2 is located in the region of G+C rich DNA upstream of ORF1. Database search revealed that ORF2 encodes a protein with high similarity to *Homo sapiens* thioesterase II (hTE, Liu et al., J. Biol. Chem., 272:13779 (1997)) and *E. coli* thioesterase II encoded by the *tesB* gene (Naggert et al., J. Biol. Chem., 266:11044 (1991)). The nucleotide sequence of ORF2 (*TES1*) is designated SEQ ID NO:57. The deduced amino acid sequence of the *TES1* protein is designated SEQ ID NO:58.

Modular structure of CPS1. Predicted CPS1 protein (1743 amino acids, M_r 193235) contains two structurally similar modules, both of which are similar to SafB1, the first module of saframycin synthetase B (overall 25% identity; 50%

similarity) and have apparent amino-acid-activating and thiolation domains but lack methyltransferase activity, thus appearing to be typical type I modules (Figure 3). The number of amino acids in each module is different: the first module (CPS1A) consists of 574 amino acids (from the first residue of core 1 to the last residue of core 6), which is larger than most type I modules; the second module (CPS1B) has 530 amino acids, which is average. The distance between the two modules is 193 amino acids, much shorter than most peptide synthetases (500-600 amino acids), but this distance is not highly conserved, i.e., an opposite variation is found in HC-toxin synthetase and cyclosporine synthetase, both of which have about 1,000 amino acids between the first and second amino-acid-activating module (see Table 6F).

Tables 6A-F show a comparative alignment of core amino acid sequences in CPS1A and CPS1B with those of other peptide synthetases. In each of Tables 6A-F, the first column shows the names of peptide synthetases; the second indicates the position of the first residue aligned in the original amino acid sequence of each protein; the last column on the right indicates the number of amino acids between two cores (6A-E, in parentheses) or the distance between two adjacent amino-acid-activating modules (Table 6F, in parentheses). The extra column in 6F, shows the total number (underlined) of residues in each amino-acid-activating module in which the aligned core sequence is located. The consensus of each core sequence is on the top, which includes identical or similar residues found in all peptide synthetases or with only a few exceptions (active site also indicated by asterisks). SafB1: the first module in saframycin Mx1 synthetase B of *Myxococcus xanthus* (Genbank Accession No. U24657); GrsA: gramicidin S synthetase A of *Bacillus brevis* (SWISS PROT Accession No. P14687); HTS1A and HTS1B: the first two modules in HC-toxin synthetase of *Cochliobolus carbonum* (Q01886); EsynA and EsynB: two modules in enniatin synthetase of *Fusarium scirpi* (EMBL Accession No. Z18755); ACVA and ACVB: the first two modules in ACV synthetase of *Aspergillus nidulans* (SWISS PROT P19787); CysnA and CsynB: the first two modules in

cyclosporine synthetase of *Tolypocladium nivenm* (EMBL Accession No. Z28383).

Table 6A: A Comparative Amino Acid Sequence Alignment of the Amino-Acid-Activating Domain (Core 1).

Consensus	X L K A G X X X V P I D P X X															SEQ ID NO: 73		
	10																	
CPS1A	165	C	F	I	A	G	V	V	A	V	P	I	N	S	V	D	(74)	SEQ ID NO: 61
CPS1B	931	C	F	V	L	G	A	V	C	I	P	M	A	P	I	D	(74)	SEQ ID NO: 62
SafB1	96	C	L	Y	A	G	V	V	A	V	P	V	Y	P	P	D	(77)	SEQ ID NO: 63
GrsA	109	V	L	K	A	G	-	G	Y	V	P	I	D	I	E	Y	(77)	SEQ ID NO: 64
HTS1A	301	I	L	K	A	G	G	V	C	V	P	I	D	P	R	Y	(82)	SEQ ID NO: 65
HTS1B	1906	V	V	Q	A	G	G	V	F	V	L	L	E	P	G	H	(80)	SEQ ID NO: 66
EsynA	556	V	L	K	A	G	H	A	F	T	L	I	D	P	S	D	(63)	SEQ ID NO: 67
EsynB	1626	I	L	K	A	N	L	A	Y	L	P	L	D	V	R	S	(65)	SEQ ID NO: 68
ACVA	361	V	W	K	S	G	A	A	Y	V	P	I	D	P	T	Y	(76)	SEQ ID NO: 69
ACVB	1455	V	W	K	S	G	G	A	Y	V	P	I	D	P	G	Y	(67)	SEQ ID NO: 70
CsynA	556	I	L	K	A	H	L	A	Y	L	P	L	D	I	N	V	(70)	SEQ ID NO: 71
CsynB	1642	I	L	K	A	G	H	A	Y	L	P	L	D	V	N	V	(68)	SEQ ID NO: 72

Table 6B: A Comparative Amino Acid Sequence Alignment of the Amino-Acid-Activating Domain (Core 2).

Consensus	F T S G X T G X P K G V X X X H R X I																				SEQ ID NO:74	
	10																					
CPS1A	253	F	S	R	A	P	T	G	D	L	R	G	V	V	L	S	H	R	T	I	(312)	SEQ ID NO:75
CPS1B	1019	W	T	Y	W	-	T	P	D	Q	R	A	V	Q	L	G	H	S	Q	I	(226)	SEQ ID NO:76
											*											
SafB1	187	Y	T	S	G	S	T	A	D	P	K	G	V	V	L	T	H	R	N	L	(213)	SEQ ID NO:77
GrsA	190	Y	T	S	G	T	T	G	N	P	K	G	T	M	L	E	H	K	G	I	(166)	SEQ ID NO:78
HTS1A	397	F	T	S	G	S	T	G	V	P	K	C	I	V	V	T	H	S	Q	I	(154)	SEQ ID NO:79
HTS1B	2000	F	T	S	G	-	T	G	V	P	K	G	A	V	A	T	H	Q	A	Y	(166)	SEQ ID NO:80
EsynA	633	F	T	S	G	S	T	G	I	P	K	G	I	M	I	E	H	R	S	F	(165)	SEQ ID NO:81
EsynB	1706	F	T	S	G	S	T	G	K	P	K	G	V	M	I	E	H	R	A	I	(169)	SEQ ID NO:82
ACVA	451	Y	T	S	G	T	T	G	F	P	K	G	I	F	K	Q	H	T	N	V	(172)	SEQ ID NO:83
ACAB	1538	Y	T	S	G	T	T	G	R	P	K	G	V	T	V	E	H	H	G	V	(181)	SEQ ID NO:84
CsynA	640	F	T	S	G	S	T	G	K	P	K	G	V	M	I	E	H	R	G	I	(172)	SEQ ID NO:85
CsynB	1724	F	T	S	G	S	T	G	K	P	K	G	V	M	I	E	H	R	G	V	(174)	SEQ ID NO:86

*An insertion (2 residues between R and A) is not shown.

Table 6C: A Comparative Amino Acid Sequence Alignment of the Amino-Acid-Activating Domain (Core 3).

Consensus	10															SEQ ID NO:87	
	G	E	L	X	V	X	G	X	G	L	A	R	G	Y			
CPS1A	583	G	E	I	W	V	D	S	P	S	L	S	G	G	F	(32)	SEQ ID NO:88
CPS1B	1209	G	E	I	W	V	Q	S	E	A	N	A	Y	S	F	(25)	SEQ ID NO:89
SafB1	418	G	E	I	W	V	R	G	P	S	V	A	Q	G	Y	(23)	SEQ ID NO:90
GrsA	374	G	E	L	C	I	G	G	E	G	L	A	R	G	Y	(23)	SEQ ID NO:91
HTS1A	569	G	E	L	L	I	E	S	G	H	L	A	D	K	Y	(31)	SEQ ID NO:92
HTS1B	2184	G	E	L	I	I	E	G	S	I	L	C	R	G	Y	(26)	SEQ ID NO:93
EsynA	816	G	E	L	V	I	E	S	A	G	I	A	R	D	Y	(30)	SEQ ID NO:94
EsynB	1893	G	E	L	V	V	T	G	D	G	V	G	R	G	Y	(32)	SEQ ID NO:95
ACVA	640	G	E	L	H	I	G	G	L	G	I	S	K	G	Y	(30)	SEQ ID NO:96
ACVB	1728	G	E	L	Y	L	G	G	E	G	V	V	R	G	Y	(30)	SEQ ID NO:97
CsynA	830	G	E	L	V	V	S	G	D	G	L	A	R	G	Y	(23)	SEQ ID NO:98
CsynB	1916	G	E	L	V	V	T	G	D	G	L	A	R	G	Y	(23)	SEQ ID NO:99

Table 6D: A Comparative Amino Acid Sequence Alignment of the Amino-Acid-Activating Domain (Core 4).

Consensus	Y - R T G D L X R										SEQ ID NO:100	
CPS1A	628	F	L	R	T	G	L	L	G	F	(13)	SEQ ID NO:101
CPS1B	1301	Y	V	R	T	G	D	L	G	F	(9)	SEQ ID NO:102
SafB1	454	W	L	R	T	G	D	L	G	F	(11)	SEQ ID NO:103
GrsA	410	Y	-	K	T	G	D	Q	A	R	(8)	SEQ ID NO:104
HTS1A	609	Y	-	R	T	G	D	L	V	R	(8)	SEQ ID NO:105
HTS1B	2223	Y	-	K	T	G	D	L	V	R	(8)	SEQ ID NO:106
EsynA	860	Y	-	R	T	G	D	L	A	C	(9)	SEQ ID NO:107
EsynB	1939	Y	-	R	T	G	D	R	M	R	(10)	SEQ ID NO:108
ACVA	684	Y	-	K	T	G	D	L	A	R	(9)	SEQ ID NO:109
ACVB	1772	Y	-	K	T	G	D	L	V	R	(11)	SEQ ID NO:110
CsynA	866	Y	-	R	T	G	D	R	A	R	(10)	SEQ ID NO:111
CsynB	1956	Y	-	R	T	G	D	R	A	R	(10)	SEQ ID NO:112

Table 6E: A Comparative Amino Acid Sequence Alignment of the Amino-Acid-Activating Domain (Core 5).

Consensus	10										20										SEQ ID NO:113			
	L	G	R	X	D	X	Q	V	K	I	R	G	X	R	I	E	L	G	E	V		E		
CPS1A	645	L	G	-	-	L	Y	E	D	R	I	R	-	Q	R	V	E	*N	G	Q	L	E	(61)	SEQ ID NO:114
GrsA	427	L	G	R	I	D	N	Q	V	K	I	R	G	H	R	V	E	L	E	E	V	E	(120)	SEQ ID NO:115
HTS1B	627	L	G	R	K	D	T	Q	V	K	M	N	G	Q	R	F	E	L	G	E	V	E	(162)	SEQ ID NO:116
HTS1A	2248	V	G	R	S	D	T	Q	I	K	L	A	G	Q	R	V	E	L	G	D	V	E	(163)	SEQ ID NO:117
EsynA	878	L	G	R	M	D	S	Q	V	K	I	R	G	Q	R	V	E	L	G	A	V	E	(139)	SEQ ID NO:118
EsynB	1958	F	G	R	M	D	N	Q	F	K	I	R	G	N	R	I	E	A	G	E	V	E	(549)	SEQ ID NO:119
ACVA	702	L	G	R	A	D	F	Q	I	K	L	R	G	I	R	I	E	P	G	E	I	E	(123)	SEQ ID NO:120
ACVB	1792	L	G	R	N	D	F	Q	V	K	I	R	G	L	R	I	E	L	G	E	I	E	(116)	SEQ ID NO:121
CsynA	884	F	G	R	M	D	Q	Q	V	K	I	R	G	H	R	I	E	P	A	E	V	E	(149)	SEQ ID NO:122
CsynB	1970	F	G	R	M	D	H	Q	V	K	V	R	G	H	R	I	E	L	A	E	V	E	(561)	SEQ ID NO:123
CPS1B	1397	L	G	S	I	G	D	T	F	E	V	N	G	L	N	H	F	S	M	D	I	E	(96)	SEQ ID NO:124
SafB1	1662	S	G	R	R	K	D	L	L	V	I	R	G	R	N	Y	Y	P	Q	D	L	E	(153)	SEQ ID NO:125

*An insertion (two amino acid) between E and N in CPS1A is not shown.

The less conserved cores 5 in CPS1B and SafB1 are indicated by arrows.

Table 6F: A Comparative Amino Acid Sequence Alignment of the Thioester Formation Domain (Core 6).

Consensus	10														SEQ ID NO:126		
	F	F	X	X	G	G	D	S	L	S	E	R	C	X			
CPS1A	726	L	D	I	P	F	L	D	S	L	S	E	R	C	<u>574</u>	(193)	SEQ ID NO:127
CPS1B	1448	R	D	P	N	G	Q	D	S	Q	M	I	T	E	<u>530</u>		SEQ ID NO:128
SafB1	645	L	P	D	L	G	L	D	S	L	A	L	V	E	<u>562</u>	(590)	SEQ ID NO:129
GrsA	567	F	Y	A	L	G	G	D	S	I	K	A	I	Q	<u>471</u>		SEQ ID NO:130
HTS1A	812	F	I	H	A	G	G	D	S	I	T	A	M	Q	<u>524</u>	(1082)	SEQ ID NO:131
HTS1B	2422	F	F	S	S	G	G	N	S	M	A	A	I	A	<u>529</u>		SEQ ID NO:132
EsynA	1040	F	F	E	M	G	G	N	S	I	I	A	I	K	<u>497</u>	(906)	SEQ ID NO:133
EsynB	2530	F	F	Q	L	G	G	H	S	L	L	A	T	K	<u>917**</u>		SEQ ID NO:134
ACVA	848	F	F	R	L	G	G	H	S	I	T	C	I	Q	<u>500</u>	(595)	SEQ ID NO:135
ACAB	1931	F	F	S	L	G	G	D	S	L	K	S	T	K	<u>489</u>		SEQ ID NO:136
CsynA	1053	F	F	D	L	G	G	H	S	L	T	A	M	K	<u>510</u>	(577)	SEQ ID NO:137
CsynB	2551	F	F	N	V	G	G	H	S	L	L	A	T	K	<u>922**</u>		SEQ ID NO:138

*Active site for 4'-phosphopantetheine binding.

**Type II modules containing a methyltransferase domain (about 400 amino acids) between cores 5 and 6. All others are type I modules without this insertion.

Amino acid alignment of the two modules of CPS1 to SafB1 indicated that these modules are highly similar to each other in both overall amino acid composition and conserved motif sequences as defined by Stachelhaus and Marahiel (Stachelhaus et al., 1995, *supra*; Marahiel, 1997, *supra*). When aligned to other bacterial or fungal peptide synthetases, CPS1 only showed local similarity to cyclosporine synthetase (Weber et al., Current Genetics, 26(2):120 (1994)) and tyrocidine synthetase A (Mootz et al., J. Bacteriol., 179(21):6843 (1997)), but when the amino acids in motif regions were aligned, an overall conservation was observed. Both CPS1A and CPS1B have all five core sequences in the amino-acid-activating domain (Table 6A-E). Cores 3 and 4 are well conserved except for the replacement of an aspartic acid residue of core 4 by a leucine in CPS1A. Cores 1, 2 and 5 show weak conservation, but similar variations are also seen in SafB1. A thiolation domain is found in both modules, which contains a highly conserved motif (core 6, Table 6F). The serine residue in this motif has been shown to be the active site for 4'-phosphopantetheine attachment (Schlumbohm et al., J. Biol. Chem., 266:23135 (1991); Stein et al., FEBS Lett., 340:39 (1994)).

The distances between the six core sequences in the two modules are also largely conserved. Two exceptions are found in the first module, which has 312 amino acids between cores 2 and 3, larger than normal (150-200); 61 between cores 5 and 6, only half of that of most peptide synthetases. SafB1 also shows distance variations at these two interval regions (Table 6B and E). In addition to amino-acid-activating and thiolation domains, CPS1 also has an integrated thioesterase domain (TE) in the carboxy-terminal end of CPS1B (Figure 12). A signature sequence GX SXG (SEQ ID NO:147), which is highly conserved in animal fatty acid thioesterase type II enzymes and several peptide synthetases, is found in this domain (Table 7).

Table 7: Comparative Alignment of Amino Acid Sequences of Active Sites of Thioesterase Domains (TE) in CPS1 with those of other Peptide Synthetases.

Consensus	X	X	X	X	G	X	S	X	G	X	X	X	A	F	E	X	SEQ ID NO:139
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
	10																
CPS1-TE	1619	V	L	R	P	G	P	S	S	G	S	E	Q	H	D	Q	A (125) SEQ ID NO:140
ACVA-TE	3621	Y	H	F	I	G	W	S	F	G	G	T	I	A	M	E	I (168) SEQ ID NO:141
GrsB-TE	4267	Y	V	L	I	G	Y	S	S	G	G	N	L	A	F	E	V (186) SEQ ID NO:142
GrsT-TE	1117	F	A	F	L	G	H	S	M	G	A	L	I	S	F	E	L (157) SEQ ID NO:143
SafA-TE	6313	L	T	L	F	G	Y	S	A	G	C	S	L	A	F	E	A (173) SEQ ID NO:144
TycC-TE	93	Y	T	L	M	G	Y	S	S	G	G	N	L	A	F	E	V (163) SEQ ID NO:145
TycF-TE	76	F	A	F	F	G	H	S	M	G	G	L	V	A	F	E	L (168) SEQ ID NO:146

ACV: ACV synthetase (SWISS PROT Accession No. P19787); GrsB: gramicidin S synthetase B (P14688); GrsT: the thioesterase encoded by grsT (P14686) in gramicidin S synthetase gene cluster; SrfA: surfactin synthetase A-3 (Q08787); TycC: tyrocidine synthetase C (Genbank Accession No. AF004835); TycF: the thioesterase encoded by tycF (AF004835) in the tyrocidine synthetase gene cluster. The highly conserved residues (GX SXG; SEQ ID NO:147) are indicated by asterisks. The number on the left of each amino acid sequence indicates the original position of the first residue; the number on the right (in parentheses) indicates the distance between the last residue shown to the end of each protein.

Sequence homology analysis of TES1 protein. The predicted TES1 protein consists of 367 amino acids (M_r 41013) amino acid alignment of TES1 to hTE, TESB and *Mycobacterium tuberculosis* TESB homolog (Philipp et al., Proc. Natl. Acad. Sci. USA, 93:3132 (1996)) showed that these proteins have an overall 40% identity and 60% similarity. A highly conserved VHS motif (putative active site) is found in the C-terminal region of TES1 at a conserved position (Figure 13). All these thioesterases have no sequence similarity with the previously identified animal type I or type II thioesterases known to be involved in the chain termination of fatty acid synthesis (Naggert et al., J. Biol. Chem., 266:11044 (1991)). Interestingly, TES1 has more homology to hTE than to two bacterial genes, suggesting that both proteins belong to a new family of eukaryotic thioesterases.

Targeted disruption of CPS1. Disruption of either CPS1A or CPS1B restored the original mutant phenotype. Ten transformants from each of four individual disruption experiments using different constructs, including the plasmid recovered from the REMI insertion site in the mutant (p214B7) and three vectors for chromosome walking (p214SNP, p118BSP and p118BCS) were purified and assayed on N-cytoplasm corn. All transformants showed the same small lesion phenotype as that of the original REMI mutant. Southern blot analysis confirmed that all transformants showing the mutant phenotype resulted from homologous integration of the transforming vector that disrupted the wild type *CPS1*. No transformants showing the wild type phenotype were obtained, presumably because of the large genomic DNA fragments (over 800 bp in all disruption experiments) on the transforming vector that resulted from high efficiency of homologous recombination and the low chance to recover transformants with ectopic integration.

Example 3

Targeted disruption of CPS1 homolog in *C. victoriae*

Methods and Materials

- Strains, growth conditions and transformation. Strains of *Cochliobolus* species and relatives used for genomic DNA hybridization are listed in Table 8. The strain HyW, a victorin-producing isolate of *C. victoriae* was recovered from storage and grown on CMX medium (Turgeon et al., Mol. Gen. Genet., 201:450 (1985)) for conidiation or on oat meal agar medium (Churchill et al., Fungal Genet. Newsl., 42A:41 (1995)) for victorin detection at 24°C under warm white lights (Sylvania Inc., Danvers, MA). Transformation was done using the *C. heterostrophus* procedure (Turgeon et al., Mol. Gen. Gene., 238:270 (1993)).

Table 8. Detection of CPS1 homologs in *Cochliobolus* spp and relatives

Strain ^a	Host ^b	<i>Eco</i> RI digest ^c	Hybridization <i>Hind</i> III digest ^d	<i>Bgl</i> III digest ^e
<i>C. heterostrophus</i>	Corn			
race T (C4)	(Turf-13)	+	5.2 3.2	4.2
race O (C5)		+	5.2 3.2	4.2
<i>C. carbonum</i>	Corn ¹			
race 1 (26R13)	(hm1hm1)	+	6.6	5.0
race 2 (Yug Y)		N	6.6	5.0
race 3 (BZ1703)*		N	6.6	5.0
<i>C. victoriae</i> (HvW)	Oats (Vb)	+	N	5.0
<i>C. sativus</i> (A20)	Grasses ²	+	3.0	N
<i>C. specifer</i> (D5-7)	Grasses ²	+	N	N
<i>C. homomorphus</i> (ATCC 13409)	Unknown	N	5.8	N
<i>C. dactyloctenii</i> (7938-9)	Unknown	N	5.9	N
<i>S. turcica</i> (NK2)	Sorghum and maize ³	+	N	N
<i>S. rostrata</i> (32197)	Weeds and bamboo ⁴	+	2.8	N
<i>B. sacchari</i>	Sugarcane ⁵			

(764-1)		+	5.4 2.5	N
(1249-10)		N	5.4 2.5	N

- a. C.= *Cochliobolus*. S.= *Setosphaeria*. B.= *Bioplaris*. The name of isolates (or lab strains) of each species are given in parentheses and those known to produce host-specific toxins are underlined. *Provided by Tsukiboshi Takao (Japan) and the isolate could be either BZ1209 or BZ1703.
- b. Genotype susceptible to the host-specific toxin-producing isolate is given in parentheses. References for hosts of those species not mentioned are as follows: 1: Welz et al., Phytopathology, 83:593 (1993); Leonard et al., Phytopathology, 80:1154 (1990) (for races 2 and 3 only). 2: Domsch et al., "Compendium of Soil Fungi, Vol. 1," New York, New York:Academic Press, pp 216-222 (1980). 3: David et al., "Fungi on Plants and Plant Products," St. Paul, Minnesota:APS Press, p. 635 (1989); Thakur et al., Plant Dis., 73:151 (1989). 4: Rao et al., Indian Bot. Rep., 6:38 (1987); Bhat et al., Curr. SCI. (BANGALORE), 58:1148 (1989). 5: Yoder, Ann. Rev. Phytopathol., 18:103 (1980).
- c. Genomic DNAs (from a previously prepared gel blot filter, Rose et al., 1996, *supra*) were probed with the 3.4 kb *CPSI* fragment cloned on p214B7. "+" indicates a strong hybridization signal. All species hybridized to a large fragment (about 23 kb).
- d. Genomic DNAs selected from a collection were probed with the *CPSI* 3.2 kb fragment cloned on p214S1. The size of fragments that hybridized to the probe is given in kb. The intensities of hybridization signals were similar to each other. N = not done.
- e. Genomic DNAs were probed with the same *CPSI* fragment as in c.
- DNA manipulations and targeted disruption of the CPS1 homolog of *C. victoriae*. Genomic DNAs for probing were prepared according to Yoder, In: Genetics of Plant Pathogenic Fungi, Vol. 6, San Diego, California:Academic Press, Sidhu, ed., pp. 93-112 (1988)), or selected from a lab DNA collection (stored at 4°C). A gel blot filter bearing known genomic DNAs was also probed. Plasmid DNA preparation, restriction enzyme digestions, gel electrophoresis, gel blot analysis were done using standard protocols (Sambrook et al., 1989, *supra*). For probing, *CPSI* fragments of *C. heterostrophus* cloned on p214B7 (3.4 kb left flank) and p214S1 (3.2 kb right flank) were prepared by restriction enzyme digestion of the plasmid DNAs followed by purification using the QIAquick Gel Extraction Kit (QIAGEN Inc., Chatsworth, CA). The plasmid p18B14, which carries the 2.3 kb *Bgl*II fragment of *CPSI* interrupted by the *hygB* cassette was linearized with *Bgl*II and introduced into HvW genome.

Transformants were purified by isolation of single conidia and genomic DNAs were digested with *Bgl*II and probed with the *CPSI* 3.2 kb fragment.

Bioassays. Pathogenicity was determined by an oat plant assay. Fungal strains were grown in individual oat meal agar medium plates (60 X 15 mm) containing hygromycin B (60 µg/ml) for 10 days at 24°C under lights. Conidia were scraped from the plates and suspended in 6 ml sterilized distilled water. One ml of conidial suspension of each strain was mixed with 60 seeds of susceptible or resistant oats. Inoculated seeds were planted in 4" X 6" pots and seedlings were allowed to grow for two weeks. Seed germination rate and symptom development were recorded at different stages (4, 6, 8 and 24 days after inoculation). Detection of victorin production using HPLC analysis was done by Alice Churchill in Dr. Vladimir Macko's lab at Boyce Thompson Institute for Plant Research.

Results

Detection of *CPSI* homologs. Genomic DNAs of 12 isolates (or lab strains) of 9 fungal species hybridized to *CPSI* (Table 8). All 6 *Cochliobolus* species, including 4 known plant pathogens (*C. carbonum*, *C. victoriae*, *C. sativus* and *C. specifer*) and 2 species with unknown hosts (*C. homomorphus* and *C. dactyloctenii*) gave hybridization signals of the same intensity as that of *C. heterostrophus CPSI* fragments. Two phytopathogenic *Setosphaeria* species and *Bioplaris sacchari*, a sugarcane pathogen gave a similar hybridization intensity.

CPSI homologs appear to be polymorphic among different species, i.e., all species gave one or two unique bands when *Bgl*II or *Hind*III digested genomic DNAs were probed (except for *C. victoriae*, which showed the same hybridization pattern as *C. carbonum*) (Table 8). Interestingly, *Eco*RI digested genomic DNAs of the same species did not show polymorphisms; all species hybridized to a large fragment (about 23 kb, Table 8), indicating the absence of an *Eco*RI site in all *CPSI* homologs as in the *C. heterostrophus* gene. In *C. heterostrophus*, a >12 kb of genomic region which includes *CPSI* (5.4 kb), *TES1* (1.1 kb) and sequence downstream of the 3' end of *CPSI* has no *Eco*RI sites. In contrast to species-dependent polymorphisms, *CPSI* homologs appear to

be highly conserved among different isolates of the same species. Both *C. heterostrophus* race T and race O hybridized to the same 4.2 kb *Bgl*II fragment (or 5.2 and 3.2 kb *Hind*III fragments); all three *C. carbonum* races hybridized to the same 5.0 kb *Bgl*II fragment (or 6.6 kb *Hind*III fragment) (Table 8) and *B. sacchari* isolates 764-1 and 1249-10 hybridized to the same *Hind*III fragments (5.4 and 2.5 kb) (Table 8).

Twenty transformants were obtained from transformation of the victorin-producing isolate HvW with *Bgl*II-linearized plasmid p118B14. Six transformants were purified and assayed for both victorin production and pathogenicity to susceptible oat plants. All transformants produced wild type levels of victorin as determined by HPLC analysis, but four of them (Tx7, Tx2, Tx5 and Tx8) showed dramatically reduced virulence in the plant assay. The seed germination rate on the eighth day after inoculation is only 13-25% for wild type and two transformants (Tx9 and Tx4), but 45-63% for the other four transformants. One day 24 after inoculation, all plants emerged from the seeds inoculated with wild type, Tx9 or Tx4 were killed but most (29-63%) from the seeds inoculated with Tx2, Tx7, Tx5 or Tx8 still survived (Table 9). Southern blot analysis confirmed that transformants showing the reduced virulence phenotype resulted from homologous integration of the transforming vector that disrupted the wild type *CPSI* homolog in *C. victoriae* genome; transformants showing the wild type phenotype resulted from ectopic integration events that left the native gene intact. All transformants remained nonpathogenic to resistant oats, indicating that disruption of the *CPSI* homolog does not affect host specificity of the fungus.

Table 9. Disease development of oat plants inoculated with *C. victoriae* transformants (Tx).

Strain ^a	No. germinated ^b			Germination Rate (%) ^c	No. survivors ^d	
	4	6	8		24	%
Control-1	28	41	45	75	75	100
Control-2	40	50	50	83	50	100
Control-3	1	7	12	20	0	0
Tx2	8	26	27	45	16	59

Strain ^a	No. germinated ^b			Germination Rate (%) ^c	No. survivors ^d	
	4	6	8		24	%
Tx4	5	15	15	25	0	0
Tx5	2	24	28	47	8	29
Tx7	14	36	38	63	24	63
Tx8	7	29	29	47	13	47
Tx9	0	3	8	13	0	0

- a. Control-1 = uninoculated susceptible oat seeds. Control-2 and Control-3 = resistant and susceptible oat seeds inoculated with wild type *C. victoriae* (isolate HvW), respectively. Six transformants were tested on both resistant and susceptible seeds, but only data for the later are shown (all transformants gave the same results as Control-2 when tested on resistant seeds). Repeat experiments gave similar results (data not shown).
- b. Sixty oat seeds were used for each strain. Emerged oat plants were counted 4, 6 and 8 days after inoculation.
- c. Calculation based on the data collected on the day 8.
- d. Recorded on day 24 after inoculation. The percentage of survivors is based on the number of plants recorded on days 8 and 24.

15 Discussion

CPS1 encodes an enzyme with an adenylation domain. A gene designated *CPS1* was cloned from the corn pathogen *C. heterostrophus* using the REMI mutagenesis procedure. Structural and functional analyses strongly suggest that *CPS1* encodes an enzyme with one or more adenylation domains, e.g., a CoA ligase. CPS1 contains two repeated functional units with a modular organization, and has a thioesterase motif (GX SXG; SEQ ID NO:147). This motif has been demonstrated to be an active site for catalyzing release of medium-chain-length (C₈₋₁₂) fatty acids in fatty acid synthases and potentially for termination of peptide chains or for repeated acyl transfer reactions because the same motif is also the characteristic of acyl transferases or acyl transfer domains (AT) of fatty acid synthases (FAS) and polyketide synthases (PKS) (Krättschmar et al., *J. Bacteriol.*, 171, 5422, (1989)).

Although similar TE domains are found in certain fungal PKSs, i.e., *Aspergillus nidulans* *pksL1* gene (Feng and Leonard, *J. Bacteriol.*, 177, 6246, (1995)) and *pksST* gene (Yu and Leonard, *J. Bacteriol.*, 117, 4792, (1995)),

CPS1 is unlikely to be a polyketide synthase because: 1) it does not show any significant similarity to known PKSs, and 2) it lacks unique functional domains found in these proteins such as the ketoacyl synthase domain (KS) and the acyl transferase domains (AT) found in the *N*-terminal region of all fungal PKSs
 5 (Yang et al., 1996, *supra*). This does not exclude the possible common evolutionary origin of CPS1 and PKSs (Stachehaus and Marahiel, 1995, *supra*).

CPS1 could be responsible for biosynthesis of an unidentified peptide phytotoxin. It is well known that several *Cochliobolus* species and related filamentous fungi produce peptide toxins. These include *C. carbonum* and *C.*
 10 *victoriae*, two species most closely related to *C. heterostrophus*. The former produces HC-toxin as mentioned above; the latter produces victorin, a chlorinated cyclized peptide. *Alternaria alternata*, a plant pathogenic species from a genus closely related to *Cochliobolus*, is also known to produce several peptide toxins such as AM-toxin, a cyclic tetradepsipeptide produced by *A.*
 15 *alternata* apple pathotype and tentoxin, a cyclic tetrapeptide produced by *A. alternata* pv. *tenuis* (Nishmura and Kohmoto, 1983). These findings have lead to the postulation that, in addition to T-toxin, *C. heterostrophus* might also produce a similar secondary metabolite, such as a hypothetical "race O" toxin (Yoder, 1981).
 20 Interestingly, a *Tox*⁺, *cpsI*⁻ mutant showed reduced virulence on T-cytoplasm corn although it produced the same amount of T-toxin as wild type race T. This is unusual because the interaction between T-toxin and the T-corn-unique URF13 protein is highly specific; the same outcomes should be expected if two strains that produce the same amount of T-toxin attack the same host, T-
 25 corn. The most likely explanation for this result is that the fungal growth *in planta* has been inhibited by the host plant and the poor growth results in reduced T-toxin production which is normal when the fungus is grown in culture. Reduced virulence on T-cytoplasm corn is due to the reduced T-toxin production as that seen in leaky *Tox*⁻ mutants. This inhibition of growth could be
 30 due to the failure of suppression of the host defense mechanism by the fungus, which is mediated by the *CPS1* controlled peptide toxin. A *cpsI*⁻ mutant that

fails to produce this “suppressor” could not be able to colonize plant tissues as vigorously as wild type does, resulting in the reduced ability to cause disease as indicated by the smaller lesion phenotype. If this turns out to be the case, *CPSI* should be considered as a general virulence factor as proposed for ennatin.

5 It is possible that *cpsI* mutants are still be able to produce a certain amount of CPS1 toxin. One probability is the gene has not been completely inactivated by insertional mutagenesis or targeted disruption. The original REMI insertion occurred at core sequence 1 of CPS1A, a region that might be not critical (function of core 1 is unknown). The second targeted site is located
10 between cores 1 and 2 of CPS1B and the third is located between cores 2 and 3 of the same module. All three insertions do not disrupt critical motifs. On the other hand, *CPSI* contains a number of in-frame start codons and some of them are located immediately downstream of these insertion sites. It is possible that each of these disruptions actually resulted in two subtranscripts, one is
15 transcribed normally from the start codon of *CPSI* and stops at the insertion site and second is transcribed near one of these in-frame ATGs downstream of the insertion site and stops at the end of *CPSI*. Both transcripts could give a truncated protein that still has enzymatic activities. But these separate enzymes might have affinities for their substrates lower than that of holoenzyme. The
20 reduced production of CPS1 toxin might be due to the CPS1 holoenzyme having been split into two fractions by the vector insertion and the resulting truncated proteins being much less active than the original polypeptide. This hypothesis can be tested by construction a *C. heterostrophus* strain in which the entire *CPSI* encoding sequence has been deleted.

25 The second possibility is the existence of multiple copies of *CPSI* in the genome. Previous studies have demonstrated that the gene encoding HC-toxin synthetase (*HTSI*) is duplicated in the genome and both copies (*HTSI-1* and *HTSI-2*) are 270 kb apart in most Tox2⁺ isolates of *C. carbonum* (Ahn and Walton, 1996, *supra*). Disruption of either copy reduced HTS1 activity but did
30 not affect HC-toxin production; when both copies were disrupted, HC-toxin production was abolished (Panaccione et al, 1992, *supra*). But in contrast to the

case of *HTSI*, gel blot analysis does not indicate the presence of a second copy of *CPSI* and disruption of *CPSI* does affect the production of the putative toxin.

It is unlikely that two genes with similar organization are in the genome. An alternative postulation is that there may be a second gene which encodes a protein with the same enzyme activity as CPS1 but does not have significant sequence homology to *CPSI*. This hypothesis is hard to test unless this gene is clustered with *CPSI* and can be recovered by chromosome walking.

In conclusion, pathogenesis by *C. heterostrophus* to corn involves at least two secondary metabolites: the T-toxin, a host specific factor which determines high virulence on a particular host, T-corn and the hypothetical CPS1 toxin, a general factor (either virulence or pathogenicity factor) which contributes to basic mechanisms underlying the disease establishment by the fungus in common host plants.

Example 4

CPSI Orthologs

As described above, *Cochliobolus heterostrophus* gene *CPSI* encodes a putative peptide synthetase that appears to be a general factor for fungal virulence to its hosts. *CPSI* has been found to be highly conserved among at least 9 fungal species belonging to 3 genera including the genus *Cochliobolus* and closely related genera *Bioplaris* and *Setosphaeria*; it has been demonstrated to be required for pathogenesis by three different plant pathogens, i.e., *C. heterostrophus* race O, race T to corn and *C. victoriae* to oats (Lu, 1998, Ph.D. thesis, Cornell University).

To further explore the role of *CPSI* in fungal pathogenesis and its conservation in other fungi, genomic DNAs of additional species of *Cochliobolus* and other closely or distantly related genera were probed with *ChCPSI* by DNA-DNA hybridization (Lu, S.-W., B.G. Turgeon and O.C. Yoder. 1999. Fungal Genetics Conference, March 1999, Pacific Grove, California). Genomic DNAs of 40 field isolates (or lab strains) representing 34 fungal species belonging to 16 genera hybridized when probed with *ChCPSI* (Figure 4).

All 16 *Cochliobolus* species, including the known plant pathogens *C. carbonum*, *C. victoriae*, *C. miyabeanus*, *C. sativus* and *C. specifer*, and five genera closely related to *Cochliobolus*, i.e., *Pyrenophora*, *Setosphaeria*, *Bipolaris*, *Stemphylium* and *Alternaria* showed hybridization intensities comparable to that of *C.*

- 5 *heterostrophus* itself (Figure 4A). DNAs of species from nine distinctly related genera, including several of economic importance (e.g., *Magnaporthe grisea*, *Fusarium graminearum*, *Gaeumannomyces graminis*) or of medical importance (e.g., *Candida albicans*) hybridized weakly to *CPS1* (Figures 4B and 4C) whereas no signal was detected in DNA of the basidiomycete *Ustilago maydis*.
- 10 Homologs of *CPS1* were further identified by polymerase chain reaction (PCR) using degenerate primers designed to conserved regions of *C. heterostrophus CPS1* (*ChCPS1*). Four *CPS1* homologs were cloned and characterized. Three of them were cloned from phytopathogenic fungi, including the wheat head scab fungus *Fusarium graminearum* (*FgCPS1*, 6003 bp, SEQ ID
- 15 NO:40), the potato early blight fungus *Alternaria solani*, (*AsCPS1*, 2369 bp, SEQ ID NO:42) and the barley net blotch fungus *Pyrenophora teres* (*PtCPS1*, 2320 bp, SEQ ID NO:44). The fourth was cloned from the human pathogenic fungus *Coccidioides immitis* (*CiCPS1*, 2435 bp SEQ ID NO:46). The complete *FgCPS1* gene was cloned using both PCR amplification and plasmid rescue
- 20 procedures preceded by targeted gene disruption of this gene in the genome. The remaining three *CPS1* homologs were partially cloned by direct PCR amplification.

- The *FgCPS1* open reading frame (5125 bp) has 50% nucleotide identity to *ChCPS1* in about 4.4 kbp of overlap. No "TATA" box-like element was
- 25 found in the 5' untranslated region, but other promoter sequences including two putative "CAAT" boxes and a "CT" motif were located upstream of the start codon (ATG). There is only one putative intron found 1508 bp upstream of the stop codon (TGA) in contrast to three in *ChCPS1*.

- A putative polyadenylation signal "AATAA" is located 62 bp
- 30 downstream of the stop codon. The predicted *FgCPS1* protein (1692 amino acids, M_r 187983 Da, SEQ ID NO:41) has 68% identity, 73% similarity to

ChCPS1 in about a 1,500 amino acid overlap that contains two structurally similar modules highly similar to those of ChCPS1 (Figure 7B). FgCPS1 has no significant similarity to ChCPS1 at the C-terminus, which is shorter and lacks the thioesterase domain seen in ChCPS1.

5 *AsCPS1* (2369 bp, SEQ ID NO:42) has 76% nucleotide identity to *ChCPS1* in the entire cloned region which contains two conserved introns. The translated AsCPS1 protein (partial) includes 758 amino acids (SEQ ID NO:43) corresponding to amino acids 511-1269 in ChCPS1 and has up to 93% identity, 95% similarity to ChCPS1 (Figure 7B).

10 *PtCPS1* (2320 bp, SEQ ID NO:44) has 78% nucleotide identity to *ChCPS1* in the entire cloned region which contains only one intron. The translated PtCPS1 protein (partial) includes 758 amino acids (SEQ ID NO:45) corresponding to amino acids 511-1269 in ChCPS1 and has 93% identity, 96% similarity to ChCPS1.

15 *CiCPS1* (2435 bp, SEQ ID NO:46) has 65% nucleotide identity to *ChCPS1* in the entire cloned region which has no introns. The translated CiCPS1 protein (partial) includes 812 amino acids (SEQ ID NO:47) corresponding to amino acids 511-1040 in ChCPS1 and has 67% identity, 80% similarity to ChCPS1 (Figure 7B). Another ortholog in *Candida* was identified
20 by Southern blot (see Figure 4).

BLAST searches using SEQ ID NO:41 (Figure 6) and SEQ ID NO:47 (Figure 7A) identified orthologs of those fungal CPS1s.

Disruption of *FsCPS1* in *F. graminearum* (= *Gibberella zeae*), the wheat head scab fungus, caused significantly reduced virulence to wheat. All *cps I*⁻
25 disruptants of *F. graminearum* showed at least 50% (when inoculated with 10⁵/ml conidia) or even 80-90% (when inoculated with 10⁴/ml conidia) reduction in ability to cause a typical "white head" symptom on the host whereas in the same conditions, ectopic transformants caused disease symptoms indistinguishable from wild type. These results suggest that *CPS1* is also
30 required for pathogenesis by fungi that are distantly related to *C. heterostrophus*,

arguing that these peptide synthetase gene homologs might control biosynthesis of a general fungal virulence factor.

Discussion

Conservation of CPS1 and taxonomy. By genomic DNA hybridization, 5 *C. heterostrophus* CPS1 homologs were found in 16 additional fungal species belonging to 5 genera. Hybridization signals for some were as strong as the *C. heterostrophus* gene, indicating that CPS1 is highly conserved among these fungi. This conservation appears to match the taxonomic relationships between these species. *Cochliobolus* (anamorph *Bipolaris*) and *Setosphaeria* (anamorph 10 *Exserohilum*) are closely related genera.

Two species, *C. victoriae* and *C. carbonum*, which are able to cross to each other and thus may not be different species (Scheffer et al., 1967; Yoder et al., 1989), showed the same hybridization pattern to CPS1. *B. sacchari*, the closest asexual relative of *C. heterostrophus*, hybridized to two HindIII 15 fragments that were only seen in *C. heterostrophus* itself, but all other species gave only one distinct polymorphic band. Phylogenetic analyses using the internal transcribed spacer (ITS) sequences and fragments of the *GPD* (vanWert and Yoder, 1992) and *MAT* genes (Turgeon et al., 1993, *supra*) also put *C. victoriae*/*C. carbonum* and *C. heterostrophus*/*B. sacchari* closest to each other 20 (Turgeon and Berbee, 1997). These results might imply that CPS1 has co-evolved with these genes.

CPS1 homologs and pathogenesis. The genera *Cochliobolus* and *Setosphaeria* include many plant pathogenic species that are commonly associated with leaf spots or blights, mainly on cultivated cereals and wild 25 grasses (Sivanesan, 1987; Alcorn, 1988). This group of phytopathogenic fungi includes both mild pathogens and severe pathogens that often produce host-specific toxins (Yoder, 1980, *supra*). One of the essential questions is whether or not the various diseases on diverse host plants caused by these fungi involve common factors or depend only on individual specific factors, such as host- 30 specific toxins.

Previous studies have shown that host-specific toxins can be critical factors for determining either virulence or host-range, but they do not account for general pathogenicity since they are produced only by certain isolates in the species and the corresponding biosynthetic genes are found only in these toxin-producing isolates (Yoder et al., 1997, *supra*). In contrast, *CPSI* homologs are found in all *Cochliobolus* and *Setosphaeria* species tested so far, suggesting they are a common factor shared by this group. Disruption of the *CPSI* homolog in the oat pathogen *C. victoriae* caused dramatically reduced virulence to victorin-susceptible oats although the transformants produced wild type levels of victorin.

This result is similar to that with *C. heterostrophus* race T, in which *cpsI* disruptants still produced wild type levels of T-toxin but showed reduced virulence on T-cytoplasm corn. These results argue strongly that host-specific toxins alone are not sufficient in determining the ultimate outcome of fungus/plant interactions and suggest that the establishment of disease by these fungi also requires CPS1, which might control a pathway for general pathogenicity.

The CPS1 gene cluster and homologs could be fungal “pathogenicity islands”. In the early 1990s, studies on pathogenesis by uropathogenic *E. coli* led to the identification of pathogenicity gene clusters, termed “pathogenicity islands” (Hecker et al., 1990; Blum et al., 1994). Subsequently, similar gene clusters were identified in additional animal or human bacterial pathogens, including *Yersinia pestis*, *Helicobacter pylori* and *Salmonella typhimurium*. These islands often contain genes for production of toxins or genes encoding proteins that are capable of interacting with host defense factors or required for type III secretion systems that deliver virulence proteins into host cells. Usually, they are found only in pathogenic strains (or species); in rare cases, they occur in nonpathogenic strains of the same species or related species (Hacker et al., 1997, *supra*).

In phytopathogenic bacteria, *hrp* gene clusters have been referred to as “pathogenicity islands” because they have several features in common with “pathogenicity islands” in animal pathogenic bacteria, i.e., they are found only in

pathogenic species (required for plant pathogenicity) and contain highly conserved genes (*hrc* genes) defining the type III protein secretion system (Alfano and Collmer, 1996; Barinaga, 1996).

In plant pathogenic fungi, genes or gene clusters with characteristics of “pathogenicity islands” have been identified from certain species, i.e., in *Nectria haematococca*, the *PDA* genes for detoxifying the pea phytoalexin and other pea pathogenicity genes (*PEP*) are located on dispensable chromosomes that are found in all isolates pathogenic to pea but usually absent in all nonpathogenic isolates (VanEtten et al., 1994; Liu et al., 1997, *supra*). In the genus *Cochliobolus*, the *Tox2* gene cluster controlling the biosynthesis of HC-toxin is found only in *C. carbonum* race 1 (pathogenic to *hm1hm1* corn) and the *Tox1* genes controlling T-toxin production are found only in *C. heterostrophus* race T (highly virulent on T-cytoplasm corn); all other races of the same species and all other fungal species tested so far lack these *Tox* genes (Ahn and Walton, 1996, *supra*; Yang et al., 1996, *supra*; Yoder et al., 1997, *supra*).

CPS1 differs in two important ways compared to these fungal “pathogenicity islands”. First, it is highly conserved among several phytopathogenic *Cochliobolus* species and relatives. Second, like certain bacterial “pathogenicity islands”, *CPS1* also has homologs in “nonpathogenic” species. *C. homomorphus* and *C. dactyloctenii*, neither of which causes disease on plants, hybridized strongly to *CPS1*. This may reflect genetic changes in the “pathogenicity island” that resulted in loss of pathogenicity. In the bacterial genus *Listeria*, which includes several human or animal pathogenic species harboring highly conserved “pathogenicity islands”, the “pathogenicity island” homolog in the nonpathogenic species (*L. seeligeri*) was found to be 'silent' due to a mutation that occurred in the promoter region of a critical regulatory gene in the cluster (Hacker et al., 1997, *supra*). These features suggest that the *CPS1* gene cluster and homologs could define a new group of fungal “pathogenicity islands”.

The origin of CPS1. It is known that the evolution of pathogenicity involves two major processes. A pathogenic microorganism could originate

from nonpathogenic progenitors by slow modifications (such as point mutations and genetic recombination) of genes that were adapted for parasitic growth on hosts or by the integration of large fragments of “alien” DNA into the genome that enable the recipient to attack particular hosts (gene horizontal transfer). The latter can occur in the recent or distant evolutionary past. Subsequent vertical transmission in the lineage (if the transferred gene is stable in the recipient genome) would result in the preserve of the gene in all species that diverged after the acquisition of the gene(s) (Scheffer, 1991; Arber, 1993; Krishnapillai, 1996; Burdon and Silk, 1997).

In the past few years, substantial evidence has become available that supports the hypothesis of gene horizontal transfer. All “pathogenicity islands” in animal pathogenic bacteria are believed to have been acquired by a horizontal transfer event (recent or past) because they usually differ in G+C content from the recipient genome and have transposable elements at the boundaries of the gene clusters (Hacker et al., 1997, *supra*). The *hrp* “pathogenicity islands” do not show a significant difference in G+C content or association with transposable elements, but they are also believed to have arisen similarly because *hrc* genes in these “pathogenicity islands” show high similarity to genes defining the type III protein secretion system found in animal pathogenic bacteria as mentioned above (Alfano and Collmer, 1996; and Barinaga, 1996).

Although *CPSI* itself has several typical fungal introns and a G+C content (51.5%) similar to most known fungal genes, genomic regions (about 1.5 kb) flanking the gene have higher G+C content (>60%). Several short G+C-rich regions are also found in the gene cluster; one of the open reading frames (ORF10) has a 63.6 % G+C content. Compared to those filamentous fungal genomes characterized so far, including *N. crassa*, *A. nidulans*, *U. maydis* (all have G+C content 51-54%, see Karlin and Mrázek, 1997, *supra*), the genomic region around *CPSI* is unusual. This might suggest that the gene cluster harboring *CPSI* came from a bacterial source (since most bacterial genes are known to have a high G+C content), but has evolved into a fungal version.

Based on these data, *CPSI* homologs may have a common ancestral gene which was acquired from a bacterial species *via* horizontal transfer and then maintained by the fungal genome *via* vertical transmission in closely related lineages.

5 In the evolution process, the genus *Cochliobolus* could also have inherited a second gene (*X*) controlling the ability to take up foreign DNA, by which its ancestor took the “alien” *CPSI*. As a result, this group of fungi is able to keep trapping genes from other organisms by additional “horizontal transfers” and giving rise to new races or even new species characterized by the ability to
10 produce unique pathogenesis factors. The direct support for this hypothesis is that both the *Tox2* locus of *C. carbonum* and the *Tox1* locus of *C. heterostrophus* are associated with large fragments of “alien” DNA (A+T-rich and highly repeated) and the same could also be true for *Tox3* controlling victorin production by *C. victoriae*, although there is yet no direct experimental evidence
15 (Ahn and Walton, 1996, *supra*; Yang et al., 1996, *supra*; Yoder et al., 1997, *supra*). In contrast to *CPSI*, these gene transfers must have occurred in the recent evolutionary past because both *Tox1* and *Tox2* loci are found only in specific isolates in the species, e.g., the acquisition of *Tox1* genes probably occurred as recently as the 1960s when race T was first identified in the field
20 (Yoder et al., 1997, *supra*).

 There are other possibilities for the evolution of *CPSI*. First, each genus mentioned above could have acquired *CPSI* independently after divergence of the lineage. But this seems less likely because this would need to happen at the same time and involve the same donor organism if the fact that the homologs
25 detected in *Cochliobolus* and *Setosphaeria* gave similar hybridization signal intensity is considered. Second, the horizontal transfer of *CPSI* could have occurred at earlier time periods such as before the divergence of Pleosporales or even the Ascomycotina. To test these hypotheses, detection of *CPSI* homologs in *Pyrenophora*, *Pleospora* and other genera must be done by either genomic
30 DNA hybridization or PCR. Based on the facts discussed here, it is not unreasonable to predict that additional *CPSI* homologs will be found in other

fungus species. Further investigation could provide a direct entry point for understanding the evolution of fungal pathogenesis to plants.

Example 5

5 Other Genes Near *Cochliobolus CPS1*

Materials and Methods

Construction of genomic library of *C. heterostrophus*. The cosmid SuperCosP1-11 (kindly provided by Dr. Thomas Hohn of Mycotoxin Research Unit USDA/ARS), which is a modification of the cosmid vector cosHyg1 (Turgeon et al., 1993, *supra*), was used for library construction. Genomic DNA of strain C4 (*Tox*⁺; *MAT*-2) was prepared as previously described (Yoder, 1988, *supra*) and purified by the equilibrium centrifugation in CsCl-ethidium bromide gradients (Sambrook, et al., 1989, *supra*). Three μ g of genomic DNA was partially digested with *Mbo*I using a test series of enzyme dilutions (1.5×10^{-4} -
10 1.25 units, New England Biolabs, Beverly, MA) at 37°C for 0.5 hour. DNA from the digestions which yielded fragments with an average size of 30 kb was pooled and then dephosphorylated with Calf Intestinal Alkaline Phosphatase (CIAP, GIBCO BRL Products, Gaithersburg, MD). Two μ g of CIAP-treated DNA was ligated into the *Bam*HI site of the cosmid vector that had been digested with *Xba*I
15 and treated with CIAP. Aliquots of the ligated molecules were packaged using Gigapack II Packaging Extract (Stratagene, La Jolla, CA) according to the manufacturer's recommendations. *E. coli* strain NM554 was transfected with the packaged phage particles and selected for ampicillin resistance. Approximately 1.6×10^5 independent ampicillin resistant colonies were obtained from two
20 experiments. Cosmid DNAs were made from 16 colonies and digested with *Hind*III and *Eco*RI respectively to confirm random insertions. Colonies were scraped from each of the original LB plus ampicillin plates and stored at -70°C in 25% glycerol (one plate of colonies/per tube).

Screening of the cosmid library. A mixture of cosmid clones from 23
30 stored tubes was diluted to 10^{-4} spread on ten LB plus ampicillin plates (150 X 15 mm) and incubated at 37°C overnight. Colonies (total about 1.2×10^4) were

transferred to Colony/Plaque Screen™ Hybridization Transfer Membrane (137 Mm discs, NEN™ Life Science Products, Boston, MA) and incubated at 37°C for 8 hours. Three replicates were made of each plate (one as master filter and two for probing). For hybridization, filters carrying colonies were lysed in 0.5 N NaOH, 1.5 M NaCl for 5 minutes, neutralized twice in 1 M Tris pH 7.4, 1.5 M NaCl for 5 minutes followed by 2 X SSC for 2 minutes. Filters were air dried 30 minutes then baked in a vacuum oven at 80°C for 1 hour. Duplicate filters were probed with ³²P labeled 3.4 and 3.2 kb fragments of the *CPSI* gene (cloned on p214B7 and p214S1, respectively) that were prepared by restriction enzyme digestion and purification using QIAquick Gel Extraction Kit (QIAGEN Inc., Chatsworth, CA). Hybridization was in 6 X SSC, 1 X BLOTTO (Sambrook et al., 1989) at 65°C overnight. Then filters were then washed twice for 15 minutes, 65°C in 2 X SSC, 0.1% SDS. Cosmid clones corresponding to positive areas were transferred from the master filters into a 96-well microtiter plate (Corning Costar, Cambridge, MA) and allowed to grow at 37°C overnight. Cells were then transferred onto membranes using a frogger, incubated and processed same as above. Positive clones were purified and re-tested by hybridization with the same probes as mentioned above. The isolated cosmid clones were mapped by probing cosmid DNA digested with several enzymes with the labeled 3.4 and 3.2 kb *CPSI* fragments separately.

DNA manipulations and sequencing. Cosmid DNA was prepared using standard protocols (Sambrook, et al., 1989, *supra*). Restriction enzyme digestions, gel electrophoresis, gel blot analysis, primer design, DNA sequencing and sequence analysis were done as described above. To facilitate sequencing, three deletion constructs were made by digestion of the original cosmid clones (Table 10) with restriction enzymes that do not cut the cosmid vector, followed by religation (Table 10). Sequencing of each cosmid clone was initiated with vector-specific and *CPSI* (or *TESI*)-specific primers. Subsequently, sequences were extended by designing new primers to the previously sequenced region (Table 11).

Results

Characterization of two overlapping cosmid clones. Two cosmid clones, C4L6582 and C4L7296, were isolated by screening the library (Table 10). Gel blot analysis indicated that both cosmid clones span the vector insertion site in the REMI mutant and contain the cloned *CPSI* and *TESI* sequences described

5 above. Sequence obtained using a primer to the region immediately flanking the insertion site is the same as that in the tagged DNA recovered from the REMI mutant, confirming that no deletions or chromosome rearrangements occurred at the tagged site. Two cosmids overlap each other in a 27.9 kb region. C4L7296 (37.2 kb) carries a 30.9 kb genomic insert which hybridized to both 3.4 kb and

10 3.2 kb *CPSI* fragments. Restriction mapping and sequencing confirmed that this insert contains the entire *TESI* sequence and most of the *CPSI* sequence (4.4 out of 5.4 kb). C4L6582 (37.7 kb) carries a 31.4 kb insert that also includes the entire *TESI* sequence but only 1.1 kb of the N-terminal encoding sequence of *CPSI*. Both inserts lack the C-terminal region of *CPSI*; their 3' end is ligated to

15 the T3 end of cloning site in SuperCosP1-11. Attempts to sequence using the T7 primer were unsuccessful, presumably because the T7 end, which is close to one of the *cos* sites on SuperCosP1-11 was disrupted during the packaging process.

Table 10. Cosmid and plasmid clones used in this study

	Clones (kb)	Length	Characteristics	Reference
5	Super- al., CosP1-11	6.9	Cosmid vector for library construction containing the 2.5 kb <i>HindIII</i> - <i>SaII</i> fragment from pH1S carrying <i>hygB</i> gene fused to <i>C. heterostrophus</i> promoter 1.	Horwitz et 1997
	pUCATPHN	4.6	Cloning vector derived from pUCATPH.	This study
10	C4L6582	37.7	A cosmid clone with a 31.4 kb insert isolated from screening the library. Includes 4.0 kb region p214B7.	This study
	C4L7296	37.2	A cosmid clone with a 30.9 kb insert isolated from screening the library. Includes 6.3 kb region p214B7+p214S1.	This study
15	p6582dH	10.9	A deletion (28.8 kb) construct derived from digestion of C4L6582 with <i>HindIII</i> .	This study
	p6582dS	21.1	A deletion (16.6 kb) construct derived from digestion of C4L6582 with <i>SacI</i> .	This study
20	p7296dX	9.0	A deletion (28.2 kb) construct derived from digestion of C4L7296 with <i>XhoI</i> .	This study
	pDXPS*	13.6	Ligation of 7296dX digested with <i>XhoI</i> to the <i>SaII</i> -digested pUCATPHN.	This study
25	pDXPSH*	6.5	A plasmid derived from pDXPS by <i>HindIII</i> digestion and religation of a 6.5 kb <i>HindIII</i> fragment containing the entire pUCATPHN sequence flanked by 1.2 kb of the 5' end of CPS1 and 0.5 kb 3' end of C4L7296 sequence	This study
30	* Designed for deletion of the 28.2 kb of genomic region (= deleted from p7296dX, including 3.6 kb <i>CPSI</i> N-terminal encoding sequence) but transformation of wild type was unsuccessful.			

Table 11. Primers used for sequencing genomic DNA on C4L7296 and C4L6582

5	Name ^a	Position ^b	Sequence ^c	Template ^d	Origin
	F-I				
	214RP7		SEQ ID NO:148	A	p214B7
	1. RP8	4940	SEQ ID NO:149	A	7296RP
	2. RP9	592	SEQ ID NO:150	A	7296RP8
10	3. RP10	4124	SEQ ID NO:151	A	7296RP9
	4. RP11	3790	SEQ ID NO:152	A	7296RP10
	5. RP12	3424	SEQ ID NO:153	A	7296RP11
	6. RP13	2970	SEQ ID NO:154	A	7296RP12
	7. RP14	2362	SEQ ID NO:155	A	7296RP13
15	8. RP15	1764	SEQ ID NO:156	A	7296RP14
	9. RP16	1169	SEQ ID NO:157	A	7296RP15
	10. RP17	647	SEQ ID NO:158	A	7296RP16
	F-II				
	214RP2		SEQ ID NO:159	B	p214B7
20	11. SRP1	3095	SEQ ID NO:160	A	6582dSRP2
	12. SRP2	2755	SEQ ID NO:161	A	7296dSRP1
	13. SRP3	2366	SEQ ID NO:162	A	7296dSRP2
	14. SRP4	2008	SEQ ID NO:163	A	7296dSRP3
	15. SRP5	1555	SEQ ID NO:164	A	7296dSRP4
25	16. SRP6	1187	SEQ ID NO:165	A	7296dSRP5
	17. SRP7	647	SEQ ID NO:166	A	7296dSRP6
	18. SFP1	3321	SEQ ID NO:167	A	6582dSRP2
	19. SFP2	3660	SEQ ID NO:168	A	7296dSFP1
	20. SFP3	3969	SEQ ID NO:169	A	7296dSFP2
30	21. SFP4	4345	SEQ ID NO:170	A	7296dSFP3
	22. SFP5	4724	SEQ ID NO:171	A	7296dsFP4
	23. SFP6	5137	SEQ ID NO:172	A	7296dSFP5
	24. SFP7	694	SEQ ID NO:173	A	7296dSFP6

F-III					
	TrpC		SEQ ID NO:174	C	pUCATPH
	214FP6		SEQ ID NO:175	D	p 214S1
	25. CFP1	463	SEQ ID NO:176	A	pDXPSTrpC
5	26. CFP2	903	SEQ ID NO:177	A	7296pUCFP1
	27. CFP3	1334	SEQ ID NO:178	A	7296pUCFP2
	28. CFP4	1910	SEQ ID NO:179	A	7296pUCFP3
	29. CFP5	2491	SEQ ID NO:180	A	7296pUCFP4
F-IV					
10	214B7RP5		SEQ ID NO:181	E	p214B7
	30. HRP1	592	SEQ ID NO:182	F	6582dHRP5
	31. HFP1	763	SEQ ID NO:183	F	6582dHRP5

^a “RP” indicates reverse primer; “FP” indicates forward primer. Primers designed to genomic DNA on the cosmid clones are numbered in order. Primers 1-10 are preceded by “7296”; 11-24 by “7296d”; 25-29 by “7296pU” and 30-31 by “6582d”.

^b Primer position corresponds to position in the genomic sequences of each fragment.

^c Each primer sequence is given in the 5' to 3' direction.

^d Cosmids or plasmids used for sequencing reactions. A = C4L7296; B = 6582dS; C = pDXPS; D = pDXPSH; E = 6582dH; F = C4L6582.

Sequencing of C4L7296. A total of 27.4 kb additional genomic sequence 5' of *TES1* was cloned. Four fragments with totaling 16.9 kb (60%) were sequenced, three of which were sequenced using C4L7296 as template. Sequencing of Fragment I (F-I, 5.3 kb) began with primer 214B7RP7 (which matches the 5' end of *TES1*), then was followed by sequencing with primers designed to previously determined sequences. Fragment II (F-II, 6.9 kb) was started using primers to sequences flanking the *SacI* site previously determined by sequencing the deletion construct 6582dS (see Table 10) and subsequently extended in both directions. Sequence of Fragment III (F-III, 3.2 kb) was obtained in a complicated manner as part of the attempt to create a deletion construct for transformation. The first part of the sequence was obtained from the clone pDXPS derived from deletion construct 7296dX (Table 10) using the

TrpC primer and the sequence was extended to the 3' end using C4L7296 as template. A 200 bp region at the 5' end of FIII was obtained from a pDXPS derived clone, pDXPSH (Table 10), using a *CPS1*-specific primer 214S1FP6.

5 Sequencing of C4L6582. This clone contains 2.8 kb additional genomic DNA extending into the region to the left end of C4L7296. The deletion clone 6582dH (Table 10) was used to initiate sequencing of Fragment IV (F-IV, 1.5 kb) using a *TES1*-specific primer 214B7RP5 followed by one step of sequence extension in both 3' and 5' direction on C4L6582.

10 Identification of open reading frames in the sequenced region. Eleven open reading frames (ORF) were identified in the four sequenced fragments (Table 12). These ORFs are all relatively small (0.3-2.3 kb). Five ORFs contain putative introns with typical fungal characteristics (Table 13). ORF12, ORF10, ORF14, ORF5 and ORF8 are transcribed in one direction; others are transcribed in the opposite direction. ORF6 and ORF7 (in F-II) overlap and are transcribed
15 in the same direction. ORF14 and ORF9 (in F-III), ORF3 and ORF8 (in F-I) also overlap but are transcribed to the opposite directions. Most ORFs have G+C content between 50-55% in the normal range for most fungal genes with the two exceptions: ORF (0.3 kb) in the 5' end of F-III has a G+C content of 63.6%; ORF14 (0.7 kb, located 1.0 kb downstream of ORF10) has a G+C content
20 56.9%. Both ORFs are located in a G+C-rich (about 58.0%) region in F-III (positions 300-800 and 1240-2040, respectively).

Database searches suggested that three ORFs (ORF3, ORF7 and ORF11) as well as *CPS1* and *TES1* encode homologs of known proteins (see below) and others encode, if anything, proteins with unknown functions (Table 12). ORF 17
25 (SEQ ID NO:48) encodes an iron reductase (SEQ ID NO:49) and ORF15 (SEQ ID NO:55) encodes a permease/MFS transporter (SEQ ID NO:56). Figure 9A shows the results of a BLAST search with SEQ ID NO:49 and Figure 10 shows the results of a BLAST search with the polypeptide encoded by SEQ ID NO:55.

Table 12. Open reading frames (ORFs) identified in sequenced genomic regions of C4L7296 and C4L6582

	Region ^a	ORF ^b	Size (kb)	No. of introns	G+C (%)	Putative Function
5						
	F-I'	<u>ORF1</u> ^d	5.4	3	51.5	<u>Peptide synthetase</u>
	F-I'	<u>ORF2</u> ^d	1.1	1	55.5	<u>Thioesterase</u>
	F-I	<u>ORF3</u>	1.8	3	50.0	<u>DNA-binding</u>
10	F-I	ORF8	0.5	0	55.2	unknown
	F-I	<u>ORF11</u>	1.9	0	52.6	CoA transferase
	F-II	ORF5	2.3	1	54.1	unknown
	F-II	ORF6	0.5	0	51.6	unknown
	F-II	<u>ORF7</u>	1.7	1	52.0	<u>Decarboxylase</u>
15	F-III	ORF9	0.7	0	54.2	unknown
	F-III	ORF10	0.3	0	63.6	unknown
	F-III	ORF13	0.8	1	53.6	unknown
	F-III	ORF14	0.7	0	56.9	unknown
20	F-IV	ORF12	1.2	1	49.2	unknown

^a F-I'= Genomic DNA

^b The positions of ORF3-ORF14 and 17 in the sequenced fragment is indicated; ORFs corresponding to known proteins are underlined.

^c The characteristics of putative introns are given in Table 12.

25 ^d Characterized as *CPSI* and *TESI*

Table 13. Characteristics of putative introns in ORFs identified in sequenced genomic regions on cosmids C4L7296 and C4L6582

	ORF	Intron	Size (bp)	Location ^a	5'Border	3'Border	Branch site
30							
	ORF3	I	64 FI	5094-5031	GTACGT	TAG	CGCTGAC
		II	46 FI	5006-4961	GTGAGT	TAG	AGCTAAG
35		III	46 FI	4477-4432	GTACGT	CAG	AGCTGAC
	ORF5	I	48 FII	3477-3524	GTATGT	TAG	TGCTAAC
	ORF7	I	114 FII	2307-2194	GTGTGC	CAG	ATCTAAC
	ORF13	I	51 FIII	2742-2692	GTGCGT	CAG	TACTGAT
	ORF12	I	47 FIV	1007-1053	GTAAGT	TAG	GATTGAC
40							
	Consensus				GT ^A / _G YGT	T ^T / _C AG	NRCTAAC ^b

^a Number of the fragment followed by the position of the first and last nucleotide of the intron with respect to the total sequence.

^b Y = Pyrimidine (T or C); R = purine; N = purine or pyrimidine.

Discussion

Two cosmids define a large gene cluster. The *C. heterostrophus CPSI* gene was cloned by identification of genomic DNA fragments recovered from the tagged site in a mutant generated using REMI insertional mutagenesis. Characterization of two overlapping cosmid clones in this study has proved that no deletions or chromosome rearrangements are associated with the gene tagging event, because both cosmids carry the same fragment which span the REMI insertion site and the nucleotide sequence in this region is the same as that of recovered genomic DNA from the tagged site. This undoubtedly clarifies the identity of *CPSI*, which is the major biosynthetic gene. Mapping and sequencing of the two cosmids extended the sequence by 27.4 kb from the previously cloned fragment, leading to the characterization of 38.7 kb of contiguous genomic DNA, the largest genomic region analyzed so far in *C. heterostrophus*. In addition to *CPSI* and *TESI*, sequence analysis of this region revealed at least 11 open reading frames; three of them, designated as *DBZ1*, *CAT1* and *DEC2*, respectively, apparently encode functional proteins (Table 13). The tight linkage of these genes suggests that they may be involved in the same pathway.

In filamentous fungi, in some cases, genes in pathways for biosynthesis of secondary metabolites are dispersed on different chromosomes, e.g., the cephalosporin C pathway genes in *Acremonium chrysogenum* (Mathison et al., 1993, *supra*) and the melanin pathway genes in *Colletotrichum lagenarium* (Kubo et al., 1996, *supra*). In other cases, tightly linked genes are usually found to be functionally related to a common pathway. This clustering organization has been exemplified by the sterigmatocystin pathway genes of *Aspergillus nidulans*, in which 25 coordinately regulated transcripts are found in a 60 kb genomic region (Brown et al., 1996) and the trichothecene pathway genes of *Fusarium sporotrichioides*, in which 9 genes are clustered in a 25 kb region and 8 of them have been shown to be required for the pathway function (Hohn et al., 1995). The genes involved in biosynthesis of certain fungal peptides are also

found as clusters. The tight linkage between *CPSI* and these additional genes might reveal the presence of a novel secondary metabolite pathway in *C. heterostrophus*. In this pathway, *CPSI* is the major structural gene since it encodes a large multifunctional enzyme with all catalytic activities required for synthesis of a secondary metabolite, presumably a peptide phytotoxin; other genes may carry out different functions required for coordinate operation of the pathway, such as regulation, posttranslational modification or substrate processing as discussed below.

Significance of the *CPSI* gene cluster. Both functional and structural analyses strongly support the hypothesis that the *CPSI* gene cluster controls a novel biosynthetic pathway. Pathway genes have been studied only in a few filamentous fungi mainly for industrial purposes (Keller et al., 1997, *supra*). For plant pathogenic fungi, little is known about pathway genes for fungal pathogenesis. In *C. heterostrophus*, recent cloning of two *Tox1* genes *PKSI* (Yang et al., 1996, *supra*) and *DECI* (Rose et al., 1996, *supra*) have contributed to a breakthrough in understanding the molecular mechanism for biosynthesis of T-toxin, a virulence determinant in the fungus/corn interaction. But further identification of related pathway genes has been unsuccessful because the two genes are located on different chromosomes and each is embedded in A+T-rich DNA (Yoder et al., 1997, *supra*). In contrast, the *CPSI* cluster provides a good opportunity to explore a pathogenesis pathway.

First, it resides in a "normal" sequence region. G+C content of a 50-55% is found in most of the cloned sequences and no A+T-rich DNA is associated with either end of the cloned region. This would facilitate cloning of additional pathway genes by further chromosome walking, by screening of cosmid libraries or the targeted integration and plasmid rescue. Second, it contains a regulatory gene (*DBZI*) which is presumably linked to a signal transduction pathway. Isolation of genes that interact with *DBZI* could reveal novel factors mediating the molecular communication between fungal pathogen and the host plant. Further characterization of *DBZI* (along with position-specific disruption or deletion) would be also helpful in determining the limit of the gene cluster,

because tightly linked genes involved in a common pathway are often coordinately regulated by the same regulatory factor (Keller et al., 1997, *supra*). Finally, *CPS1* genes are found in both race T and race O, and its homologs are also found in other *Cochliobolus* species. Presence of high G+C content may
5 imply that these genes evolved from a bacterial ancestor and the conservation in these fungi may correlate with the phytopathogenic function of the gene products encoded by the *CPS1* cluster. Further investigation of this cluster should provide insights into the evolution of general pathogenicity factors among this group of fungi.

10 ORF17 is an iron reductase (SEQ ID NO:49) and ORF15 is a permease/MFS transporter (SEQ ID NO:56). Ferric reductases are a group of enzymes found in bacteria, fungi, plants and animals that are responsible for reduction of ferric iron to ferrous iron, an absorptive form used by the organism. They have been well studied in *S. cerevisiae*, *C. albicans* and *H. capsulatum* and
15 the like. The yeast FER1 has been expressed in tobacco (Oki et al., 1999).

Previous studies have shown that FER genes could be important pathogenic determinants. Timmerman and Woods have proposed that in *H. capsulatum* FER could play critical roles in the acquisition of iron in three different ways: from inorganic or organic ferric salts, from host Fe(III) binding
20 proteins (transferrin and the like), and from siderophores produced by the fungus itself (to reduce and release the iron chelated by the siderophore molecules).

On the other hand, iron sequestration in response to microbial infection has been demonstrated to be a host defense mechanism. The infection-related iron acquisition system in the pathogen can be considered to be an important
25 mechanism against host defense and for a successful colonization by the pathogen in the host cells. This could be a general mechanism for all pathogenic fungi.

CPS1 may encode an enzyme which is responsible for biosynthesis of a novel siderophore with unusual amino acid, hydroxyl acid and architecture. The
30 CPS1 siderophore can compete with the host for iron acquisition when the fungus enters its host cells where the iron is limited due to host sequestration. In

particular, for root pathogens such as *C. victoriae*, sequestration may be stronger in the root surface. This could explain why the *cps1* mutant showed drastically reduced virulence. The FER1 could be required to release iron from the CPS1 siderophore which explains its location near the *CPS1* gene. Moreover, fungal
5 strains could be cultured in iron-limiting conditions because CPS1, and likely other genes in the cluster maybe turned on only during conditions of iron depletion.

All publications, patents and patent applications are incorporated herein
10 by reference. While in the foregoing specification, this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details herein may be varied considerably without departing from
15 the basic principles of the invention.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a fungal nucleic acid segment
which encodes a polypeptide which is substantially similar to a
5 polypeptide encoded by a nucleic acid sequence comprising an open
reading frame comprising SEQ ID NO:46, SEQ ID NO:48, or SEQ ID
NO:55, or the complement thereof.
2. An isolated polynucleotide comprising a fungal nucleic acid segment
10 which is substantially similar to a nucleic acid sequence comprising an
open reading frame comprising SEQ ID NO:46, SEQ ID NO:48, SEQ ID
NO:55, or the complement thereof.
3. An isolated polynucleotide comprising a fungal nucleic acid segment
15 which hybridizes under stringent hybridization conditions to SEQ ID
NO:46, SEQ ID NO:48, SEQ ID NO:55, or the complement thereof.
4. The isolated polynucleotide of claim 1, 2 or 3 which consists of SEQ ID
NO:46, SEQ ID NO:48 or SEQ ID NO:55 of the complement thereof.
20
5. The isolated polynucleotide of claim 1, 2 or 3 wherein the nucleic acid
segment is from *Ascomycota*.
6. The isolated polynucleotide of claim 1, 2 or 3 wherein the nucleic acid
25 segment is from a pathogenic fungus.
7. The isolated polynucleotide of claim 1 wherein the nucleic acid segment
encodes a polypeptide having at least 80% identity to a polypeptide
comprising SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56.
30

8. The isolated polynucleotide of claim 1 wherein the nucleic acid segment encodes a polypeptide having at least 90% identity to a polypeptide comprising SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:56.
- 5 9. An isolated polypeptide encoded by the polynucleotide of any one of claims 1 to 8.
10. An expression cassette comprising a promoter operably linked to the polynucleotide of any one of claims 1 to 8.
- 10 11. A recombinant vector comprising the polynucleotide of any one of claims 1 to 8 wherein the vector is capable of being stably transformed into a host cell.
- 15 12. The vector of claim 11 wherein the polynucleotide is operably linked to a promoter operable in a eukaryotic host cell.
13. The expression cassette of claim 10 or vector of claim 11 wherein the polynucleotide is in sense orientation.
- 20 14. The expression cassette of claim 10 or vector of claim 11 wherein the polynucleotide is in antisense orientation.
15. The vector of claim 11 wherein the polynucleotide is operably linked to a promoter operable in a prokaryotic host cell.
- 25 16. A host cell comprising the expression cassette of claim 10.
17. A host cell comprising the vector of claim 11.
- 30

18. The host cell of claim 16 or 17 which is selected from the group consisting of bacteria, yeast, plant and mammal.
19. A method for identifying an agent having fungicidal or mycocidal activity, comprising:
- 5 a) contacting a fungus with an agent that binds to the polypeptide of claim 9; and
- b) identifying an agent having fungicidal or mycocidal activity.
- 10 20. An agent identified by the method of claim 19.
21. A method for identifying an inhibitor of a polypeptide, comprising:
- a) contacting a host cell which expresses a polypeptide encoded by the polynucleotide of any one of claims 1 to 8 with an agent; and
- 15 b) identifying an agent that inhibits the activity of the polypeptide.
22. An agent identified by the method of claim 21.
23. A method of inhibiting the growth or pathogenicity of a fungus, comprising contacting the fungus with the agent of claim 20 or 22 in an amount sufficient to inhibit the growth or pathogenicity of the fungus.
- 20 24. A method for identifying an agent having fungicidal or mycocidal activity, comprising:
- 25 a) contacting a fungus with an agent that inhibits the activity of the polypeptide of claim 9; and
- b) identifying an agent having fungicidal or mycocidal activity.
25. A method for identifying an agent that modulates a polypeptide associated with pathogenicity of a fungus, comprising:
- 30

- a) contacting a fungus with an agent that binds the polypeptide of claim 9; and
- b) identifying an agent that modulates the pathogenicity of the fungus.
- 5
26. A method for identifying an agent that modulates the pathogenicity of a fungus, comprising:
- a) contacting a fungus with an agent that inhibits the activity of the polypeptide of claim 9; and
- 10 b) identifying an agent that modulates the pathogenicity of the fungus
27. A method of identifying agents that alter the phenotype of a fungal pathogen or mycogen, comprising:
- 15 a) contacting an agent to be tested with one or more cells of a fungal pathogen or mycogen which comprises a nucleotide sequence encoding a polypeptide that is substantially similar to SEQ ID NO:47, SEQ ID NO:49, or SEQ ID NO:56; and
- 20 b) detecting or determining whether the agent selectively modulates expression or function or metabolic pathways associated with the polypeptide, thereby altering a phenotype of the cells relative to cells not contacted with the agent.
28. The method of claim 27 wherein the polypeptide is associated with virulence or pathogenicity.
- 25
29. The method of claim 27 wherein the agent alters the activity of the polypeptide.
- 30 30. The method of claim 27 further comprising identifying an agent having fungicidal, mycocidal or anti-pathogenic activity.

31. The method of claim 27 wherein cellular growth is detected or determined.
- 5 32. The method of claim 27 wherein the activity of the polypeptide is detected or determined.
33. The method of claim 27 wherein virulence is detected or determined.
- 10 34. The method of claim 27 wherein the pathogen expresses the polypeptide.
35. The method of claim 27 wherein the pathogen does not express the polypeptide.
- 15 36. A method of identifying agents that alter the phenotype of a fungal pathogen or mycogen, comprising
- a) contacting an agent to be tested with one or more cells of a fungal pathogen or mycogen wherein the cells have a mutation in a nucleic acid sequence corresponding to the polynucleotide according to any one of claims 1 to 8 which mutation results in overexpression or underexpression of the encoded polypeptide;
- 20 b) detecting or determining whether the agent selectively modulates expression or function or metabolic pathways associated with the polypeptide, thereby altering a phenotype of the cells relative to one or more wild type cells not contacted with the agent.
- 25 37. The method of claim 27 or 36 wherein the pathway is associated with the production of a toxin or siderophore.
- 30 38. The method of claim 27 or 36 wherein the pathway is associated with iron metabolism, uptake or absorption.

39. The method of claim 27 or 36 wherein the pathway is associated with growth, virulence or pathogenicity.
40. An isolated antibody which specifically binds to the polypeptide of claim 9.
41. The antibody of claim 40 which is a monoclonal antibody.
42. The antibody of claim 40 which is a polyclonal antibody.
43. The method of claim 19, 23, 24, 25, 26, 27 or 36 wherein the fungus is a recombinant fungus.
44. The method of claim 43 wherein the fungus comprises a recombinant DNA molecule which encodes the polypeptide.
45. The method of claim 44 wherein the recombinant DNA molecule is overexpressed.
46. The method of claim 44 wherein the fungus comprises an antisense recombinant DNA molecule for the polypeptide.
47. The method of claim 44 wherein the genome of the fungus is disrupted so that the endogenous gene which encodes the polypeptide is not expressed.
48. A therapeutic method comprising: administering to an animal suspected of being infected with a fungal pathogen an effective amount of the agent of claim 19 or 22.

49. A method to prevent or inhibit infection of an animal or plant by a fungal pathogen, comprising: administering to the animal or plant an effective amount of the agent of claim 19 or 22 for a time and under conditions sufficient to inhibit or prevent fungal growth or reproduction.
- 5
50. The method of claim 51 or 52 wherein the animal is a human.
51. The method of claim 51 or 52 wherein the agent is topically administered.
- 10
52. A nucleic acid sequence of a polynucleotide of any one of claims 1 to 8.
53. The nucleic acid sequence of claim 52 which is stored on a computer readable medium.
- 15
54. An amino acid sequence of a polypeptide of claim 9.
55. The amino acid sequence of claim 54 which is stored on a computer readable medium.
- 20
56. The method of claim 48 or 49 wherein the animal is immunocompromised.
57. The method of claim 48 or 49 wherein the animal has
- 25 Coccidioidomycosis.
58. The method of claim 48 or 49 wherein the animal is subjected to immunosuppressive therapy.
- 30
59. The method of claim 48 or 49 wherein fungal iron metabolism is inhibited.

60. The method of claim 49 wherein the agent is administered to a plant.
61. The method of claim 60 wherein the agent is administered by spraying.
- 5
62. A transformed plant, the genome of which expresses a chimeric DNA molecule which encodes a gene product which confers resistance or tolerance to the plant to a fungal pathogen by inhibiting fungal iron metabolism or siderophore production.
- 10

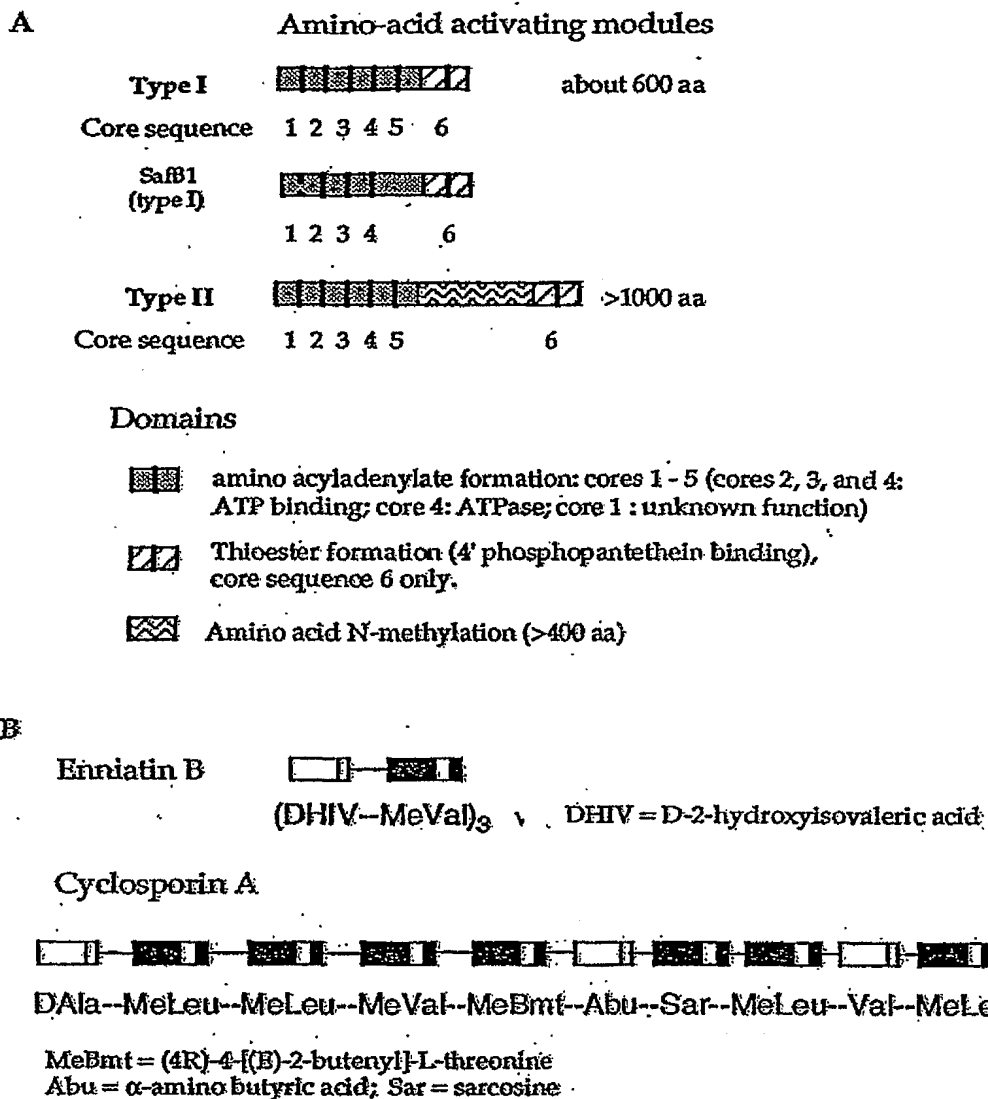


Figure 1

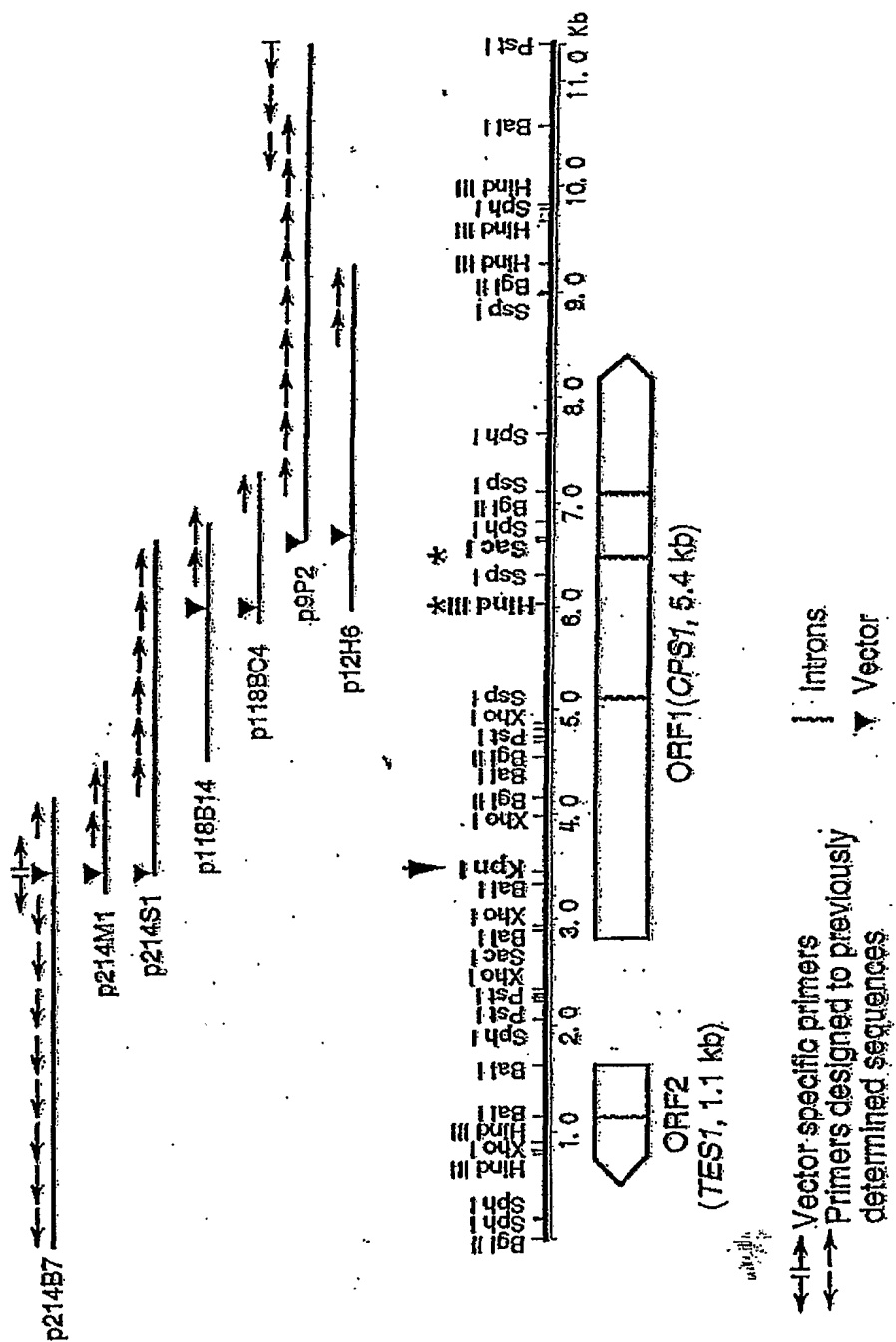


Figure 2

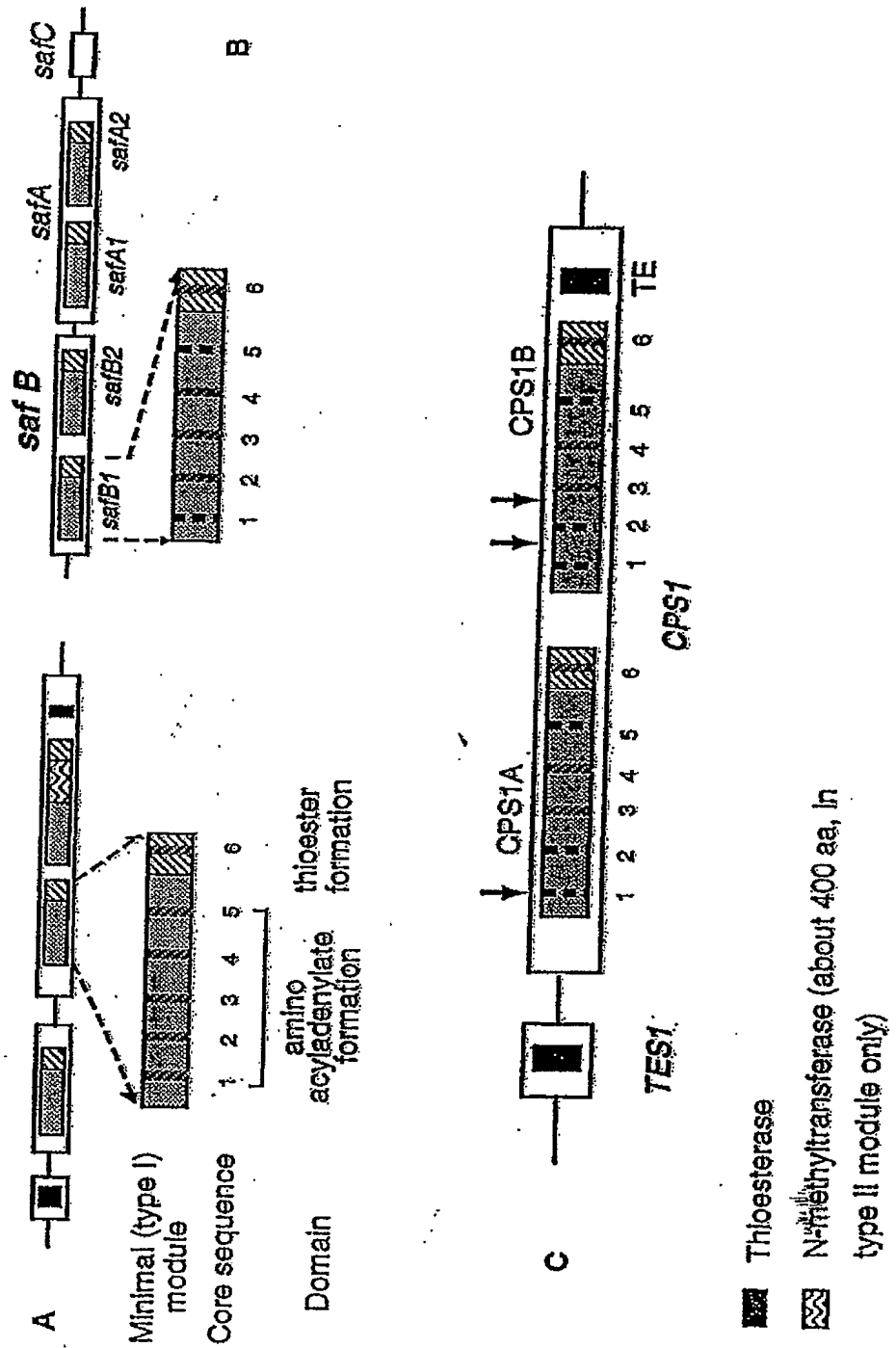
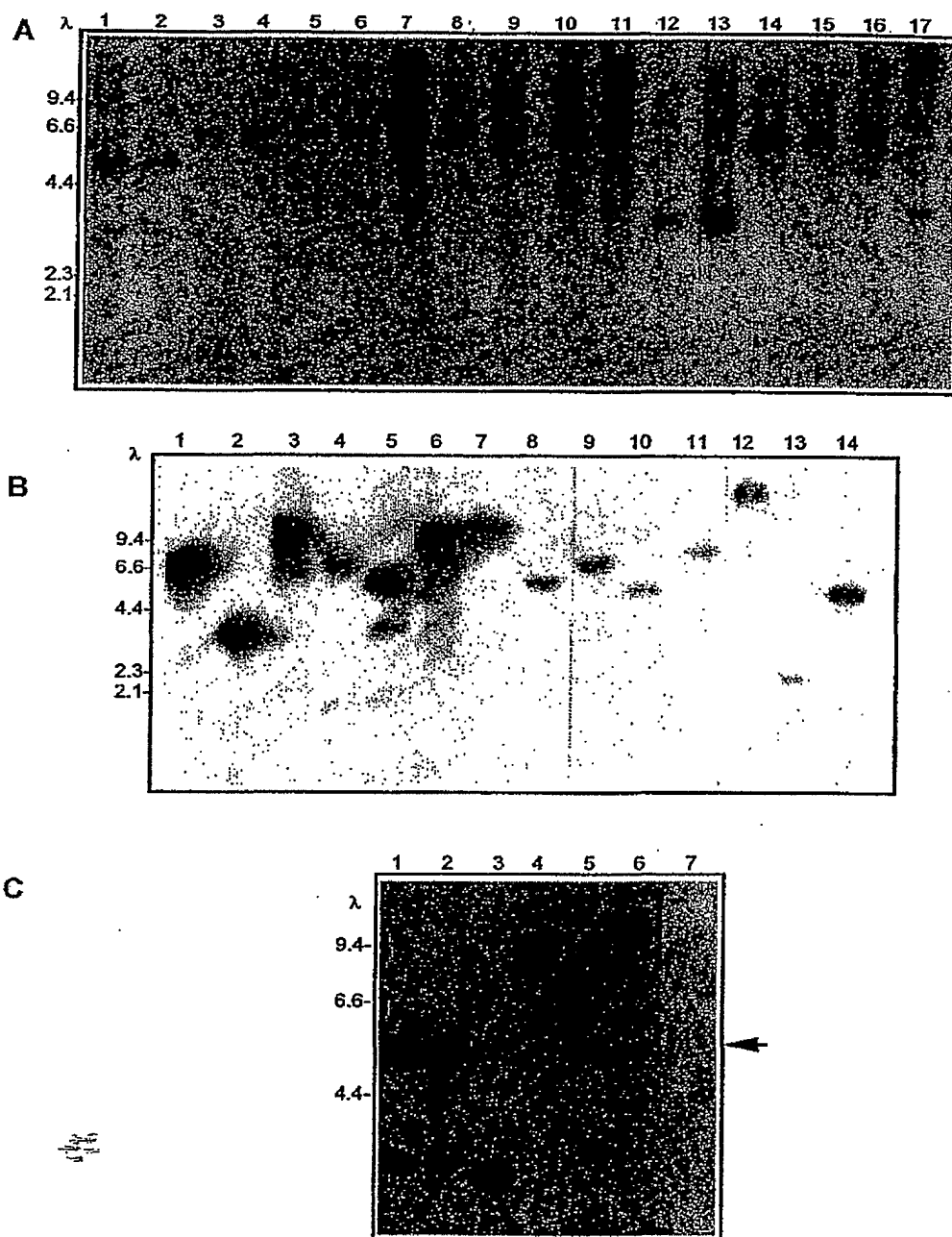


Figure 3A-C



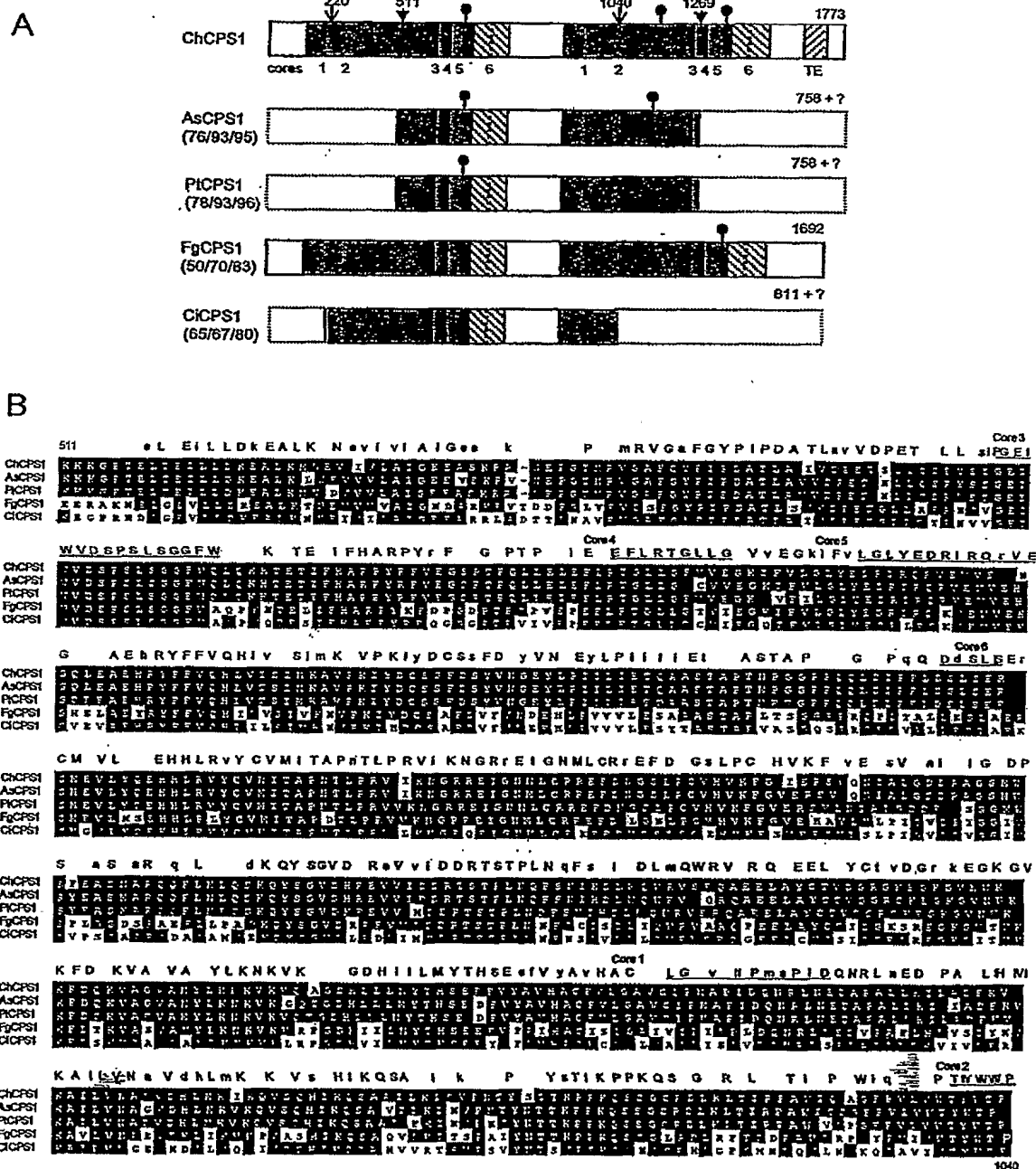
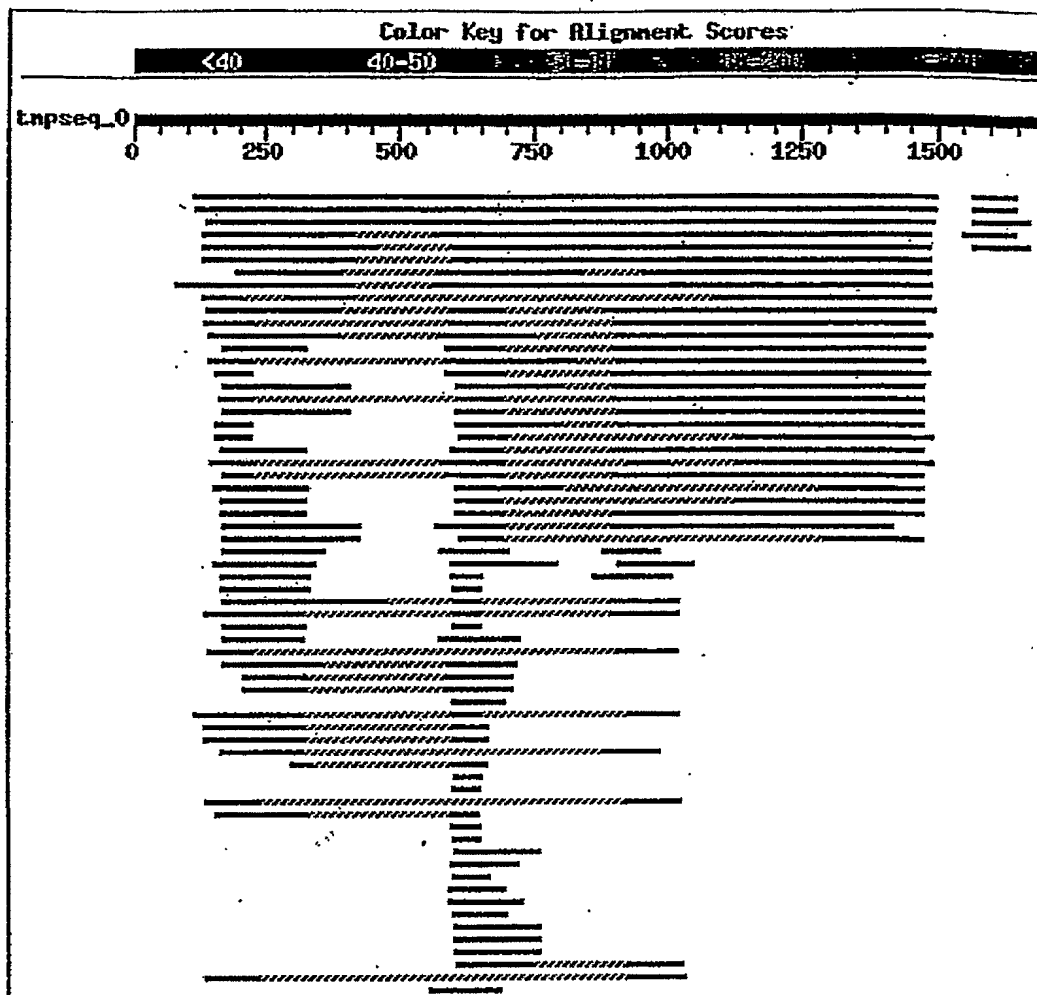


Figure 5A-B



Sequences producing significant alignments:

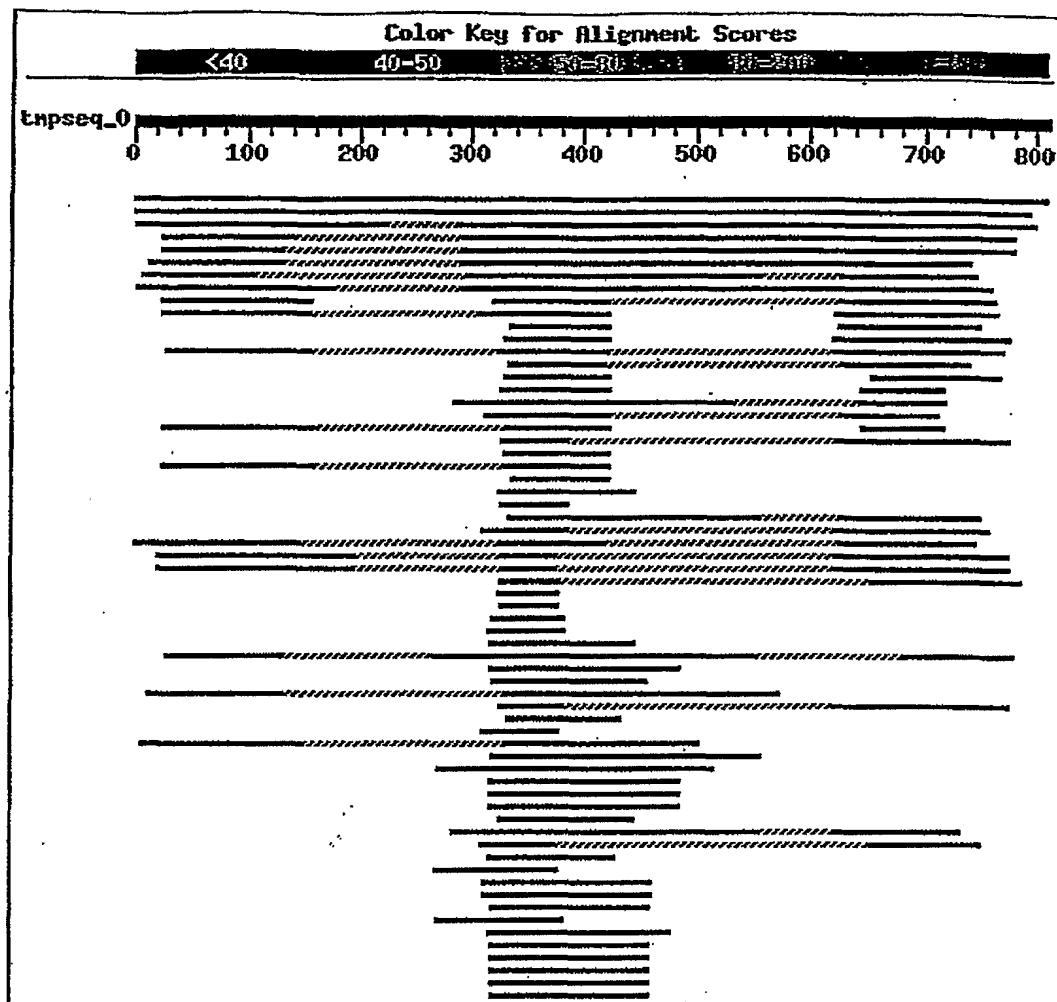
Score E
(bits) Value

sp Q10250 YD22 SCHPO HYPOTHETICAL 170.7 KD PROTEIN C56F8.02...	1407	0.0
sp Q09773 YA84 SCHPO HYPOTHETICAL 162.4 KD PROTEIN C22F3.04...	576	e-163
ref NP_014736.1 Yor093cp [Saccharomyces cerevisiae] >gi 21...	561	e-158
gb AAF64300.1 (AF246991) unknown [Drosophila melanogaster]	249	2e-64
sp Q14689 Y184 HUMAN HYPOTHETICAL PROTEIN KIAA0184 >gi 1136...	247	6e-64
sp Q9Y2E4 Y934 HUMAN HYPOTHETICAL PROTEIN KIAA0934 >gi 4589...	247	6e-64
dbj BAA95987.F (AB040896) KIAA1463 protein [Homo sapiens]	212	3e-53
pir T34061 hypothetical protein F28B3.1 - Caenorhabditis e...	205	5e-51
gb AAF47364.1 (AE003467) CG7020 gene product [Drosophila m...	150	1e-34
pir T34918 polyketide synthase - Streptomyces coelicolor >...	125	4e-27
gb AAG02359.1 AF210249 18 (AF210249) peptide synthetase NRP...	106	2e-21
pir T18551 saframycin Mx1 synthetase B - Myxococcus xanthu...	106	2e-21
sp Q10976 YT30 MYCTU HYPOTHETICAL 67.9 KDA PROTEIN RV2930 >...	87	2e-15
gb AAG05812.1 AE004669 9 (AE004669) probable non-ribosomal ...	86	3e-15
pir A70635 probable fadD31 protein - Mycobacterium tubercu...	85	8e-15
pir C70669 probable acyl-CoA synthetase (EC 6.2.1.7) - Myc...	82	7e-14
pir F70522 probable polyketide synthase - Mycobacterium tu...	77	2e-12

Figure 6

pir B70668	probable Acyl-CoA Synthetase - Mycobacterium tu...	74	1e-11
pir C70634	probable fadD30 protein - Mycobacterium tubercu...	73	2e-11
qb AAB52538.1	(U75685) acyl-CoA synthase [Mycobacterium bo...]	73	2e-11
emb CAB36629.1	(AL035480) putative acyl-CoA synthase [Myco...]	71	8e-11
pir T31307	type I fatty acid synthase homolog - Cryptospor...	71	1e-10
sp Q50586 YF21 MYCTU	HYPOTHETICAL 63.1 KDA PROTEIN RV1521 >...	69	3e-10
pir S73072	u0002r protein - Mycobacterium tuberculosis >gi...	69	4e-10
pir B70820	probable polyketide synthase MTCY409 - Mycobact...	68	7e-10
pir A70877	probable acyl-CoA synthase - Mycobacterium tuber...	62	4e-08
pir E70887	probable fadD32 protein - Mycobacterium tubercu...	60	3e-07
pir S72716	4-coumarate--CoA ligase homolog - Mycobacterium...	59	4e-07
qb AAC83455.1	(AF117694) malonyl CoA synthetase [Rhizobium...]	54	1e-05
sp P94547 LCFA BACSU	LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	52	7e-05
pir H72454	probable long-chain-fatty-acid--CoA ligase APE2...	50	2e-04
pir T03221	probable polyketide synthase module 1 - Strepto...	50	2e-04
sp P29212 LCFA ECOLI	LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	49	4e-04
qb AAF86393.1 AF235504_14	(AF235504) FkbB [Streptomyces hyg...]	48	6e-04
sp P46450 LCFA HAEIN	LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	48	6e-04
sp P25464 ACVS CEPAC	DELTA-(L-ALPHA-AMINOADIPYL)-L-CYSTEINY...	48	8e-04
pir E69378	long-chain-fatty-acid--CoA ligase (fadD-5) homo...	48	8e-04
qb AAB09715.1	(U12891) ORF5 [Pseudomonas aeruginosa]	48	0.001
qb AAF00957.1 AF183408_5	(AF183408) McyG [Microcystis aerug...]	47	0.001
sp O30408 TYCB BACBR	TYROCIDINE SYNTHETASE II [INCLUDES: AT...]	47	0.001
db BAB12213.1	(AB032549) peptide synthetase and polyketid...	47	0.001
qb AAF28840.1 AF118888_1	(AF118888) malonyl CoA synthetase ...	47	0.001
pir A61209	hypertension-associated protein SA - rat	46	0.002
qb AAF95133.1	(AE004273) long-chain-fatty-acid--CoA ligase...	46	0.003
emb CAA06324.1	(AJ005061) LchAB protein [Bacillus lichenif...]	45	0.006
qb AAD04758.1	(U95370) lichenysin synthetase B; LicB [Baci...]	45	0.007
qb AAF41909.1	(AE002506) long-chain-fatty-acid--CoA ligase...	45	0.007
emb CAB84971.1	(AL162757) long-chain-fatty-acid--CoA-ligas...	45	0.007
qb AAF08797.1 AF184956_4	(AF184956) MycC [Bacillus subtilis]	45	0.010
pir S19560	proline-rich protein MP4 - mouse >gi 53182 emb ...	44	0.013
db BAB06823.1	(AP001517) long-chain fatty-acid-CoA ligase...	44	0.013
pir T07943	probable AMP-binding protein - rape >gi 1617272...	44	0.013
pir T18841	hypothetical protein C01G6.7 - Caenorhabditis e...	43	0.022
qb AAF02529.1 AF150669_1	(AF150669) long-chain-fatty-acid-C...	43	0.028
emb CAB38084.1	(AJ006977) Tal [Myxococcus xanthus]	43	0.028
db BAA37141.1	(AB022340) SA [Mus musculus]	43	0.037
pir T17428	FK506 polyketide synthase - Streptomyces sp. (s...	43	0.037
db BAB10742.1	(AB023035) 4-coumarate-CoA ligase-like prot...	43	0.037
pir H69545	probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	43	0.037
pir T30226	polyketide synthase - Streptomyces hygroscopicu...	42	0.049
pir T17463	rifamycin polyketide synthase modules 1-3 - Amy...	42	0.049
ref NP_058566.1	SA rat hypertension-associated homolog [Mu...]	42	0.049
pir YGBSG1	phenylalanine racemase (ATP-hydrolyzing) (EC 5....)	42	0.049
db BAB01855.1	(AP000377) long-chain-fatty-acid CoA ligase...	42	0.064
pir T05038	4-coumarate--CoA ligase homolog F13C5.180 - Ara...	42	0.064
sp O68006 BACA-BACLI	BACITRACIN SYNTHETASE 1 (BA1) [INCLUDE...	41	0.083
pir H69354	probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	41	0.11
pir B48013	proline-rich proteoglycan 2 precursor, parotid ...	41	0.11
pir B69768	probable acid--CoA ligase (EC 6.2.1.-) ydaB - B...	41	0.11
pir H71401	probable A6 anther-specific protein - Arabidops...	41	0.11
qb AAF00961.1 AF183408_9	(AF183408) McyB [Microcystis aerug...]	41	0.11
qb AAB92395.1	(U97078) microcystin synthetase B [Microcyst...]	41	0.11
emb CAA78044.1	(Z12000) AngR protein [Vibrio anguillarum]	41	0.11
pir T36248	CDA peptide synthetase I - Streptomyces coelico...	41	0.11
sp P19828 ANGR VIBAN	ANGR PROTEIN >gi 68686 pir YGVCAR ang...	41	0.11
qb AAF26925.1 AF210843_22	(AF210843) nonribosomal peptide s...	41	0.11
pir D69187	probable acid--CoA ligase (EC 6.2.1.-) MTH657 -...	41	0.11

<u>pir</u>	<u> T07908</u>	4-coumarate--CoA ligase (EC 6.2.1.12) 2 - weste...	<u>41</u>	0.11
<u>pir</u>	<u> E69438</u>	probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	<u>41</u>	0.14
<u>pir</u>	<u> E69274</u>	acetyl-CoA synthetase (acs-1) homolog - Archaeo...	<u>41</u>	0.14
<u>pdb</u>	<u> 1AMU B</u>	Chain B, Phenylalanine Activating Domain Of Gram...	<u>41</u>	0.14
<u>pir</u>	<u> T28932</u>	probable 4-coumarate--CoA ligase (EC 6.2.1.12) ...	<u>41</u>	0.14
<u>sp</u>	<u> P14687 GRSA_BACBR</u>	GRAMICIDIN S SYNTHETASE I [INCLUDES: A...	<u>41</u>	0.14
<u>qb</u>	<u> AAA58718.1 </u>	(M29703) grsA-encoded protein [Brevibacillus...	<u>41</u>	0.14
<u>qb</u>	<u> AAF42473.1 AF204401_1</u>	(AF204401) actinomycin synthetase ...	<u>40</u>	0.19
<u>qb</u>	<u> AAG06687.1 AE004752_3</u>	(AE004752) long-chain-fatty-acid--...	<u>40</u>	0.19
<u>qb</u>	<u> AAG01008.1 AF288210_1</u>	(AF288210) TRK-fused protein TFG [...	<u>40</u>	0.19
<u>sp</u>	<u> P26046 ACVT_PENCH</u>	DELTA-(L-ALPHA-AMINOADIPYL)-L-CYSTEINY...	<u>40</u>	0.19
<u>sp</u>	<u> P19787 ACVS_PENCH</u>	DELTA-(L-ALPHA-AMINOADIPYL)-L-CYSTEINY...	<u>40</u>	0.19
<u>pir</u>	<u> JX0340</u>	gramicidin S synthase 2 - Bacillus brevis >gi 5...	<u>40</u>	0.19
<u>emb</u>	<u> CAB72227.1 </u>	(A1138854) putative peptide synthetase; wit...	<u>40</u>	0.19
<u>pir</u>	<u> S73073</u>	pks002a protein - Mycobacterium tuberculosis >g...	<u>40</u>	0.25
<u>dbj</u>	<u> BAA83993.1 </u>	(AB019578) mcyB [Microcystis aeruginosa]	<u>40</u>	0.25
<u>ref</u>	<u> NP_036764.1 </u>	proline-rich protein, salivary [Rattus nor...	<u>40</u>	0.25
<u>pir</u>	<u> A70669</u>	probable acyl-CoA synthetase (EC 2.3.1.-) - Myc...	<u>40</u>	0.25
<u>qb</u>	<u> AAF79612.1 AC027665_13</u>	(AC027665) F5M15.18 [Arabidopsis ...	<u>40</u>	0.25
<u>qb</u>	<u> AAF08796.1 AF184956_3</u>	(AF184956) MycB [Bacillus subtilis]	<u>40</u>	0.25
<u>pir</u>	<u> T14591</u>	actinomycin synthetase II - Streptomyces chryso...	<u>40</u>	0.25
<u>pir</u>	<u> H69371</u>	probable acid--CoA ligase (EC 6.2.1.-) AF0976 -...	<u>39</u>	0.32
<u>qb</u>	<u> AAF46366.1 </u>	(AE003443) CG10555 gene product [Drosophila ...	<u>39</u>	0.32



Sequences producing significant alignments:

	Score	E
	(bits)	Value
sp Q10250 YD22 SCHPO HYPOTHETICAL 170.7 KD PROTEIN C56F8.02...	819	0.0
ref NP_014736.1 Yor093cp [Saccharomyces cerevisiae] >gi 21...	277	3e-73
sp Q09773 YA84 SCHPO HYPOTHETICAL 162.4 KD PROTEIN C22F3.04...	169	1e-40
sp Q14689 Y184 HUMAN HYPOTHETICAL PROTEIN KIAA0184 >gi 1136...	89	3e-16
sp Q9Y2E4 Y934 HUMAN HYPOTHETICAL PROTEIN KIAA0934 >gi 4589...	86	1e-15
gb AAF64300.1 (AF246991) unknown [Drosophila melanogaster]	81	1e-14
pir T18551 saframycin Mx1 synthetase B - Myxococcus xanthu...	80	1e-13
pir T34061 hypothetical protein F28B3.1 - Caenorhabditis e...	73	1e-11
sp Q10976 YT30 MYCTU HYPOTHETICAL 67.9 KDA PROTEIN RV2930 >...	72	2e-11
pir C70669 probable acyl-CoA synthetase (EC 6.2.1.-) - Myc...	71	6e-11
pir S73072 u0002r protein - Mycobacterium tuberculosis >gi...	68	3e-10
pir F70522 probable polyketide synthase - Mycobacterium tu...	66	1e-09
emb CAB36629.1 (AL035480) putative acyl-CoA synthase [Myco...	65	4e-09
gb AAG02359.1 (AF210249) peptide synthetase NRP...	63	1e-08
pir B70820 probable polyketide synthase MTCY409 - Mycobact...	62	2e-08
pir A70877 probable acyl-coAsynthase - Mycobacterium tuber...	61	5e-08
pir A70635 probable fadD31 protein - Mycobacterium tubercu...	59	2e-07

Figure 7A

<u>pir</u> <u>T31307</u> type I fatty acid synthase homolog - Cryptospor...	<u>59</u>	<u>2e-07</u>
<u>sp</u> <u>Q50586</u> <u>YF21 MYCTU</u> HYPOTHETICAL 63.1 KDA PROTEIN RV1521 >...	<u>59</u>	<u>2e-07</u>
<u>qb</u> <u>AAF95133.1</u> (AE004273) long-chain-fatty-acid--CoA ligase...	<u>59</u>	<u>2e-07</u>
<u>pir</u> <u>B70668</u> probable Acyl-CoA Synthetase - Mycobacterium tu...	<u>59</u>	<u>2e-07</u>
<u>qb</u> <u>AAB52538.1</u> (U75685) acyl-CoA synthase [Mycobacterium bo...	<u>59</u>	<u>3e-07</u>
<u>pir</u> <u>S72716</u> 4-coumarate--CoA ligase homolog - Mycobacterium...	<u>57</u>	<u>6e-07</u>
<u>pir</u> <u>E69378</u> long-chain-fatty-acid--CoA ligase (fadD-5) homo...	<u>54</u>	<u>7e-06</u>
<u>qb</u> <u>AAC83455.1</u> (AF117694) malonyl CoA synthetase [Rhizobium...	<u>54</u>	<u>7e-06</u>
<u>qb</u> <u>AAG05812.1</u> <u>AE004669 9</u> (AE004669) probable non-ribosomal ...	<u>53</u>	<u>1e-05</u>
<u>emb</u> <u>CAA70871.1</u> (Y09700) rpfB [Xanthomonas campestris]	<u>52</u>	<u>2e-05</u>
<u>pir</u> <u>T34918</u> polyketide synthase - Streptomyces coelicolor >...	<u>52</u>	<u>4e-05</u>
<u>sp</u> <u>P29212</u> <u>LCFA ECOLI</u> LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	<u>52</u>	<u>4e-05</u>
<u>sp</u> <u>P46450</u> <u>LCFA HAEIN</u> LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	<u>52</u>	<u>4e-05</u>
<u>qb</u> <u>AAB09715.1</u> (U12891) ORF5 [Pseudomonas aeruginosa]	<u>50</u>	<u>1e-04</u>
<u>pir</u> <u>A45062</u> long-chain-fatty-acid--CoA ligase (EC 6.2.1.3) ...	<u>49</u>	<u>2e-04</u>
<u>qb</u> <u>AAF83100.1</u> <u>AE003882 2</u> (AE003882) regulator of pathogenic...	<u>49</u>	<u>2e-04</u>
<u>pir</u> <u>H72454</u> probable long-chain-fatty-acid--CoA ligase APE2...	<u>49</u>	<u>2e-04</u>
<u>qb</u> <u>AAF28840.1</u> <u>AF118888 1</u> (AF118888) malonyl CoA synthetase ...	<u>49</u>	<u>2e-04</u>
<u>pir</u> <u>T03221</u> probable polyketide synthase module 1 - Strepto...	<u>48</u>	<u>4e-04</u>
<u>qb</u> <u>AAF86393.1</u> <u>AF235504 14</u> (AF235504) FkbB [Streptomyces hyg...	<u>48</u>	<u>5e-04</u>
<u>qb</u> <u>AAF79612.1</u> <u>AC027665 13</u> (AC027665) F5M15.18 [Arabidopsis ...	<u>47</u>	<u>7e-04</u>
<u>pir</u> <u>S73071</u> u0002q protein - Mycobacterium tuberculosis >gi...	<u>47</u>	<u>7e-04</u>
<u>dbj</u> <u>BAA95987.1</u> (AB040896) KIAA1463 protein [Homo sapiens]	<u>46</u>	<u>0.001</u>
<u>pir</u> <u>YGBSG1</u> phenylalanine racemase (ATP-hydrolyzing) (EC 5....	<u>46</u>	<u>0.001</u>
<u>pir</u> <u>T07908</u> 4-coumarate--CoA ligase (EC 6.2.1.12) 2 - weste...	<u>46</u>	<u>0.002</u>
<u>qb</u> <u>AAF47364.1</u> (AE003467) CG7020 gene product [Drosophila m...	<u>45</u>	<u>0.003</u>
<u>qb</u> <u>AAF41909.1</u> (AE002506) long-chain-fatty-acid--CoA ligase...	<u>45</u>	<u>0.003</u>
<u>emb</u> <u>CAB84971.1</u> (AL162757) long-chain-fatty-acid--CoA-ligas...	<u>45</u>	<u>0.003</u>
<u>qb</u> <u>AAD39590.1</u> <u>AC007858 4</u> (AC007858) This gene is a member o...	<u>45</u>	<u>0.005</u>
<u>dbj</u> <u>BAB06823.1</u> (AF001517) long-chain fatty-acid-CoA ligase...	<u>44</u>	<u>0.006</u>
<u>pir</u> <u>C70634</u> probable fadD30 protein - Mycobacterium tubercu...	<u>44</u>	<u>0.008</u>
<u>qb</u> <u>AAF08795.1</u> <u>AF184956 2</u> (AF184956) MycA [Bacillus subtilis]	<u>44</u>	<u>0.008</u>
<u>pir</u> <u>E69438</u> probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	<u>43</u>	<u>0.010</u>
<u>pdb</u> <u>1AMU</u> <u>B</u> Chain B, Phenylalanine Activating Domain Of Gram...	<u>43</u>	<u>0.013</u>
<u>pir</u> <u>T18841</u> hypothetical protein C01G6.7 - Caenorhabditis e...	<u>43</u>	<u>0.013</u>
<u>sp</u> <u>P14687</u> <u>GRSA BACBR</u> GRAMICIDIN S SYNTHETASE I [INCLUDES: A...	<u>43</u>	<u>0.013</u>
<u>qb</u> <u>AAA58718.1</u> (M29703) grsA-encoded protein [Brevibacillus...	<u>43</u>	<u>0.013</u>
<u>sp</u> <u>P94547</u> <u>LCFA BACSU</u> LONG-CHAIN-FATTY-ACID--COA LIGASE (LON...	<u>42</u>	<u>0.023</u>
<u>pir</u> <u>E70887</u> probable fadD32 protein - Mycobacterium tubercu...	<u>42</u>	<u>0.023</u>
<u>qb</u> <u>AAF79611.1</u> <u>AC027665 12</u> (AC027665) F5M15.17 [Arabidopsis ...	<u>42</u>	<u>0.030</u>
<u>pir</u> <u>T17463</u> rifamycin polyketide synthase modules 1-3 - Amy...	<u>42</u>	<u>0.030</u>
<u>pir</u> <u>C69471</u> probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	<u>42</u>	<u>0.030</u>
<u>sp</u> <u>Q01886</u> <u>HTS1 COCCA</u> HC-TOXIN SYNTHETASE (HTS) >gi 167219 g...	<u>41</u>	<u>0.039</u>
<u>pir</u> <u>A45086</u> HC-toxin synthetase - fungus (Cochliobolus carb...	<u>41</u>	<u>0.039</u>
<u>pir</u> <u>T07943</u> probable AMP-binding protein - rape >gi 1617272...	<u>41</u>	<u>0.039</u>
<u>qb</u> <u>AAD40664.1</u> <u>AF150686 1</u> (AF150686) 4-coumarate:coenzyme A ...	<u>41</u>	<u>0.039</u>
<u>sp</u> <u>P31685</u> <u>4CL2 SOLTU</u> 4-COUMARATE--COA LIGASE 2 (4CL) >gi 10...	<u>41</u>	<u>0.051</u>
<u>pir</u> <u>T30226</u> polyketide synthase - Streptomyces hygroscopicu...	<u>41</u>	<u>0.051</u>
<u>sp</u> <u>P14912</u> <u>4CL1 PETCR</u> 4-COUMARATE--COA LIGASE 1 (4CL) >gi 82...	<u>41</u>	<u>0.051</u>
<u>qb</u> <u>AAD34542.1</u> <u>AF139644 1</u> (AF139644) luciferase [Phrixothrix...	<u>41</u>	<u>0.051</u>
<u>pir</u> <u>T28932</u> probable 4-coumarate--CoA ligase (EC 6.2.1.12) ...	<u>41</u>	<u>0.051</u>
<u>sp</u> <u>P31684</u> <u>4CL1 SOLTU</u> 4-COUMARATE--COA LIGASE 1 (4CL) >gi 10...	<u>41</u>	<u>0.051</u>
<u>pir</u> <u>T09755</u> 4-coumarate--CoA ligase (EC 6.2.1.12) 4CL2 - lo...	<u>41</u>	<u>0.051</u>
<u>pir</u> <u>T05038</u> 4-coumarate--CoA ligase homolog F13C5.180 - Ara...	<u>41</u>	<u>0.051</u>
<u>sp</u> <u>P14913</u> <u>4CL2 PETCR</u> 4-COUMARATE--COA LIGASE 2 (4CL) >gi 28...	<u>41</u>	<u>0.051</u>
<u>qb</u> <u>AAF37732.1</u> <u>AF052221 1</u> (AF052221) 4-coumarate--CoA ligase...	<u>41</u>	<u>0.067</u>
<u>qb</u> <u>AAF37733.1</u> <u>AF052222 1</u> (AF052222) 4-coumarate--CoA ligase...	<u>41</u>	<u>0.067</u>
<u>pir</u> <u>T03390</u> 4-coumarate--CoA ligase (EC 6.2.1.12) isoform 2...	<u>41</u>	<u>0.067</u>
<u>pir</u> <u>T17428</u> FK506 polyketide synthase - Streptomyces sp. (s...	<u>41</u>	<u>0.067</u>
<u>emb</u> <u>CAB81058.1</u> (AL161502) 4-coumarate--CoA ligase-like pro...	<u>41</u>	<u>0.067</u>

<u>pir</u>	<u> T07909</u>	4-coumarate--CoA ligase (EC 6.2.1.12) 1 - weste...	<u>40</u>	<u>0.088</u>
<u>gb</u>	<u> AAB18637.1</u>	(U50845) 4-coumarate:coenzyme A ligase [Nico...	<u>40</u>	<u>0.088</u>
<u>gb</u>	<u> AAF02529.1</u>	<u> AF150669 1</u> (AF150669) long-chain-fatty-acid-C...	<u>40</u>	<u>0.088</u>
<u>gb</u>	<u> AAF91310.1</u>	<u> AF239687 1</u> (AF239687) 4-coumarate:CoA ligase ...	<u>40</u>	<u>0.088</u>
<u>gb</u>	<u> AAG29784.1</u>	<u> AF235050 7</u> (AF235050) putative ligase [Strept...	<u>40</u>	<u>0.12</u>
<u>gb</u>	<u> AAG06688.1</u>	<u> AE004752 4</u> (AE004752) long-chain-fatty-acid--...	<u>40</u>	<u>0.12</u>
<u>pir</u>	<u> C75364</u>	long-chain fatty acid--CoA ligase - Deinococcus...	<u>39</u>	<u>0.15</u>
<u>pir</u>	<u> H69354</u>	probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	<u>39</u>	<u>0.15</u>
<u>sp</u>	<u> O30408 TYCB BACBR</u>	TYROCIDINE SYNTHETASE II [INCLUDES: AT...	<u>39</u>	<u>0.15</u>
<u>pir</u>	<u> H69545</u>	probable fatty-acid--CoA ligase (EC 6.2.1.-) fa...	<u>39</u>	<u>0.15</u>
<u>dbj</u>	<u> BAA05006.1</u>	(D25416) luciferase [Photuris pennsylvanica]	<u>39</u>	<u>0.20</u>
<u>pir</u>	<u> A70551</u>	probable acid--CoA ligase (EC 6.2.1.-) fadD35 -...	<u>39</u>	<u>0.20</u>
<u>pir</u>	<u> S73073</u>	pks002a protein - Mycobacterium tuberculosis >g...	<u>39</u>	<u>0.26</u>
<u>sp</u>	<u> P17814 4CL ORYSA</u>	4-COUMARATE--COA LIGASE >gi 82454 pir ...	<u>39</u>	<u>0.26</u>
<u>emb</u>	<u> CAB84715.1</u>	(AL162756) putative acyl-CoA ligase [Neisse...	<u>39</u>	<u>0.26</u>
<u>pir</u>	<u> H69371</u>	probable acid--CoA ligase (EC 6.2.1.-) AF0976 -...	<u>39</u>	<u>0.26</u>
<u>pir</u>	<u> A70669</u>	probable acyl-CoA synthetase (EC 2.3.1.-) - Myc...	<u>39</u>	<u>0.26</u>
<u>emb</u>	<u> CAA53230.1</u>	(X75542) 4-coumarate:CoA ligase [Vanilla pl...	<u>39</u>	<u>0.26</u>
<u>dbj</u>	<u> BAA05005.1</u>	(D25415) luciferase [Photuris pennsylvanica]	<u>39</u>	<u>0.26</u>
<u>gb</u>	<u> AAF41652.1</u>	(AE002476) long-chain-fatty-acid--CoA ligase...	<u>39</u>	<u>0.26</u>
<u>dbj</u>	<u> BAB01715.1</u>	(AB023045) 4-coumarate:CoA ligase [Arabidop...	<u>39</u>	<u>0.26</u>
<u>pir</u>	<u> T36202</u>	probable fatty acid--CoA ligase - Streptomyces ...	<u>38</u>	<u>0.34</u>
<u>pir</u>	<u> H71401</u>	probable A6 anther-specific protein - Arabidops...	<u>38</u>	<u>0.34</u>

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

		Majority									
		10	20	30	40	50	60	70			
1	MLEVN	-----QGYFSDFTGQQMQDNRD									
1		-----AsolaniCPS1pro.PRO									
1		-----pteresCPS1pro.PRO									
1	MMSGDYAFRPEQQTGYESQ	-----FGcps1pro.PRO									
1		-----C.immitis CPS1pro.PRO									
		Majority									
		80	90	100	110	120	130	140			
23	S--YGGPN-RYSSGDAFSPTAA	-----cps1pro.PRO									
1		-----AsolaniCPS1pro.PRO									
1		-----pteresCPS1pro.PRO									
71	TVEYVGPOQRYSSDAFSPTAAMAPMLTTNDLPPPEALEYQLPLDPREVPFAIQDPHDDSTPMSKFDNI	-----C.immitis CPS1pro.PRO									
1		-----C.immitis CPS1pro.PRO									
		Majority									
		150	160	170	180	190	200	210			
89	GAVLRHRSRTQPRTTAFWVLD	-----cps1pro.PRO									
1		-----AsolaniCPS1pro.PRO									
1		-----pteresCPS1pro.PRO									
141	AAVLRHGRGRTIAKKPAYWVLD	-----FGcps1pro.PRO									
1		-----C.immitis CPS1pro.PRO									

Figure 7B

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

	220	230	240	250	260	270	280	Majority
159	VVALMGCFIAGVAVPINSVDDYQKLILLTTTQAHIALTTDNNLKAFHRDISQNRLLKWP	CGE	230	240	250	260	270	280
1	-----	-----	-----	-----	-----	-----	-----	cpslpro.PRO
1	-----	-----	-----	-----	-----	-----	-----	AsolaniCPSlpro.PRO
211	AIALLGCFIAGVAVPINDHQQYQRLNHILTTTQAHIALTTDNNLKAFORDITTQKLTPK	CGE	230	240	250	260	270	280
1	-----	-----	-----	-----	-----	-----	-----	pteresCPSlpro.PRO
	-----	-----	-----	-----	-----	-----	-----	Fgcpslpro.PRO
	-----	-----	-----	-----	-----	-----	-----	C.immitis CPSlpro.PRO
	-----	-----	-----	-----	-----	-----	-----	Majority
229	FGSYHPKKKDD-PAL-VPDLAYIEFSRAPTGDLRGVLSHRTIMHQM	CGE	230	240	250	260	270	280
1	-----	-----	-----	-----	-----	-----	-----	cpslpro.PRO
1	-----	-----	-----	-----	-----	-----	-----	AsolaniCPSlpro.PRO
281	FGSYHPKKKEDVPALVVPDLAYIEFSRAPTGDLRGVLSHRTIMHQM	CGE	230	240	250	260	270	280
10	-----	-----	-----	-----	-----	-----	-----	pteresCPSlpro.PRO
	-----	-----	-----	-----	-----	-----	-----	Fgcpslpro.PRO
	-----	-----	-----	-----	-----	-----	-----	C.immitis CPSlpro.PRO
	-----	-----	-----	-----	-----	-----	-----	Majority
299	LRDAECKFVAPAPSRNPTEVILTYLDPRESAGLILSVLEFVVGHTTVWLE	CGE	230	240	250	260	270	280
1	-----	-----	-----	-----	-----	-----	-----	cpslpro.PRO
1	-----	-----	-----	-----	-----	-----	-----	AsolaniCPSlpro.PRO
350	LRDKNGRLIGGAS---	CGE	230	240	250	260	270	280
71	-----	-----	-----	-----	-----	-----	-----	pteresCPSlpro.PRO
	-----	-----	-----	-----	-----	-----	-----	Fgcpslpro.PRO
	-----	-----	-----	-----	-----	-----	-----	C.immitis CPSlpro.PRO

Alignment Report of Unfilled, using Clustal method with PAM250 residue weight table.

-LL-ADYPGLKRAAYNYQQDPMATRNFKN-EPNF-S-KLCLIDTLITVDCEHFEILLADRWLRPLRNPRAR Majority										
	430	440	450	460	470	480	490			
369	NILADYPGLKRAAYNYQQDPMATRNFKNTEPNFASVKICLIDTLITVDCEHFEILLGDRYFRPLRNPRAR									
1	-----									
1	-----									
417	TIMTADYPGLKRAAYNYQQDPMATRNFKNTEPNFQMTIKLCLIDTLITVDGSGHEVILADRWLRPLRNPRAR									
135	ALLAADYPGLKRAAYNYQQDPMATRNFKNSEPNFESSLKCLIDTLITVDCEHFEILLADRWLRPLRNPRAR									

DPETGLLCSFYSIGEIWVDSPLSGGFNQLQKHTETIFHARPYRFEVGSPTQLLEFLRTGLLGFVVE Majority

GKIFVLGLYEDRIRQREVEVHEGQLEAEHRYFFVQHLVTSIMKAVPKIYDCSSFDSYVNGEYLPILLIET Majority

QAASTAPTNPGGPPQQQLDIPFLDSLSERCMEVLYQEHHLRVYCVMITAPNTLPRVVKGRREIGNMLCRR Majority

15/34

Alignment Report of Untitled. usha Clustal method with PAM250 residue weight table.

	850	860	870	880	890	900	910	Majority
	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
780	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
270	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
270	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
828	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
552	EFDNGSLPCVHVKFG	TERSQNI	ALGDDP	AGCMWS	EASWARQO	FLMLQDK	QYSGVDHRE	VIDDRTSTP
	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
850	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
340	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
340	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
898	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
622	LNQFSNIHDL	MQWRVSRQAE	ELAYCTV	DGRGKEG	KGVNWK	FFDQKVAG	VAMYLKNK	VKVRPGD
	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
920	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
410	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
410	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
968	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV
692	HSEEFVYAV	HACFVLGAV	CIPMAP	IDQNR	LNE	DAPALL	HILAD	EKVKAILV

Col 1

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK Majority										
	1060	1070	1080	1090	1100	1110	1120	color		
990	SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK									
480	SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK									
480	SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK									
1038	SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK									
762	SAAILKISVPNTYNTTKPPKQSSGCRDLKLTIRPAWIOGFFVLVWYTWTPDQRRRIAVQLGHSQIMALCK									
	C.immitis CPS1pro.PRO									
	Majority									
	1130	1140	1150	1160	1170	1180	1190			
1060	VQKETQMTSTRPVLGCVRSTIGLGFHTCLMGIFLAAPTIVSPVDFEAQNPNIFFQTLTRYKIKDAYAT									
550	VQKETQMTSTRPVLGCVRSTIGLGFHTCLMGIFLAAPTIVSPVDFEAQNPNIFFQTLTRYKIKDAYAT									
550	VQKETQMTSTRPVLGCVRSTIGLGFHTCLMGIFLAAPTIVSPVDFEAQNPNIFFQTLTRYKIKDAYAT									
1108	VQKETQMTSTRPVLGCVRSTIGLGFHTCLMGIFLAAPTIVSPVDFEAQNPNIFFQTLTRYKIKDAYAT									
811	VQKETQMTSTRPVLGCVRSTIGLGFHTCLMGIFLAAPTIVSPVDFEAQNPNIFFQTLTRYKIKDAYAT									
	C.immitis CPS1pro.PRO									
	Majority									
	1200	1210	1220	1230	1240	1250	1260			
1130	SQMLDHAIAARGAGKNMALHELKKNLMIATDGRPRVDVYQVRVHFAPASLDRTAINTVYSHVLNPMVASRS									
620	SQMLDHAIAARGAGKNMALHELKKNLMIATDGRPRVDVYQVRVHFAPASLDRTAINTVYSHVLNPMVASRS									
620	SQMLDHAIAARGAGKNMALHELKKNLMIATDGRPRVDVYQVRVHFAPASLDRTAINTVYSHVLNPMVASRS									
1178	SQMLDHAIAARGAGKNMALHELKKNLMIATDGRPRVDVYQVRVHFAPASLDRTAINTVYSHVLNPMVASRS									
811	SQMLDHAIAARGAGKNMALHELKKNLMIATDGRPRVDVYQVRVHFAPASLDRTAINTVYSHVLNPMVASRS									
	C.immitis CPS1pro.PRO									

Alignment Report of Untitled. using Clustal method with PAM250 residue weight table.

YMCIEPIELHLDVXALRRGLVMPVDPDTEPGALLVQDSGMVFPVSTQIAIVNPETNQLCLVGEYGEIIV-S Majority									
	1270	1280	1290	1300	1310	1320	1330		
1200	YMCIEPIELHLDVXALRRGLVMPVDPDTEPNALLVQDSGMVFPVSTQISIVNPETNQLCLVGEYGEIIVQS	core 3 cpslpro.PRO							
690	YMCIEPIELHLDVXALRRGLVMPVDPDTEPGALLVQDSGMVFPVSTQISIVNPETNQLCLVGEYGEIIV--	AsolaniCPSlpro.PRO							
690	YMCIEPIELHLDVXALRRGLVMPVDPDTEPGALLVQDSGMVFPVSTQIAIVNPETNQLCLVGEYGEIIV--	pteresCPSlpro.PRO							
1248	YMCIEPIELHLDVXALRRGLVMPVDPDTEPNALLVQDSGMVFPVSTQIAIINPESRIHCLDGEYGEIIVDS	Fgcpslpro.PRO							
811	C.immitis CPSlpro.PRO								
EA---SFY-SK---DAERF-GR--DGDPN--Y-RTGDLGFLH-V-RPIGPNGA-VDMQVLFVLG-IG-TF Majority									
	1340	1350	1360	1370	1380	1390	1400		
1270	EANAYSFYMSKERLDAERFNGRTIDGDPNVRYVRTGDLGFLHVSRTPIGPNGAPVDMQVLFVLGSGIDTF	core 4 cpslpro.PRO							
758	AsolaniCPSlpro.PRO								
758	pteresCPSlpro.PRO								
1318	EACVKSFYCSKDAERFDGRALDGDPNICVIRTGDLGFLHVSRTPIGPNGAQVDMQVLFVLGSGIDTF	Fgcpslpro.PRO							
811	C.immitis CPSlpro.PRO								
E-NGL-HF-MDIE-SVE-CHRNIV--GCAVFOAGGLVV-VE--R---LAS-VPVIVNAILNEHQ---DI Majority									
	1410	1420	1430	1440	1450	1460	1470		
1340	EVNGLNHFSMDIEQSVRCHRNIVPGCCAVFOAGGLVVVVEIFRNFLASMPVIVNAILNEHQIIVDI	cpslpro.PRO							
758	AsolaniCPSlpro.PRO								
758	pteresCPSlpro.PRO								
1388	EVNGLNHFSMDIEQSVRCHRNIVPGCCAVFOAGGLVVVVEIFRNFLASMPVIVNAILNEHQIIVDI	Fgcpslpro.PRO							
811	C.immitis CPSlpro.PRO								

V-V-FV-KGDF-RSRLGEKQKGKIL-GWV-RK-RT-AQ-SIRD	P-RASM	Majority
---	--------	----------

--S--G--AP-----N-----Q-----M-P-----P-----EL Majority

--S--G--AP-----N-----Q-----M-P-----P-----EL Majority

D-D-TT-HS-P-G-P-Q-P-Majority

D-D-TT-HS-P-G-P-Q-P-Majority

Sequencing strategy

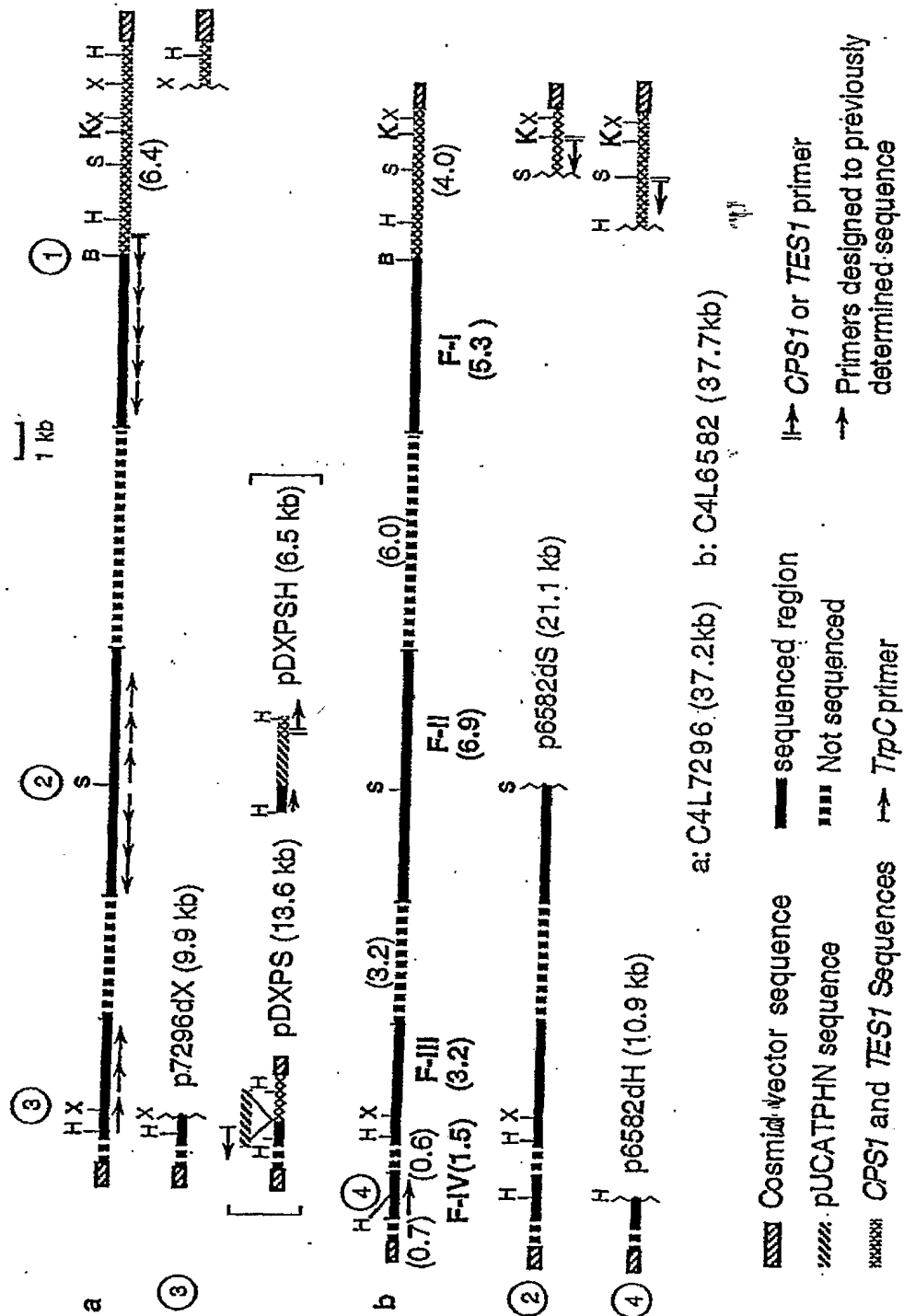
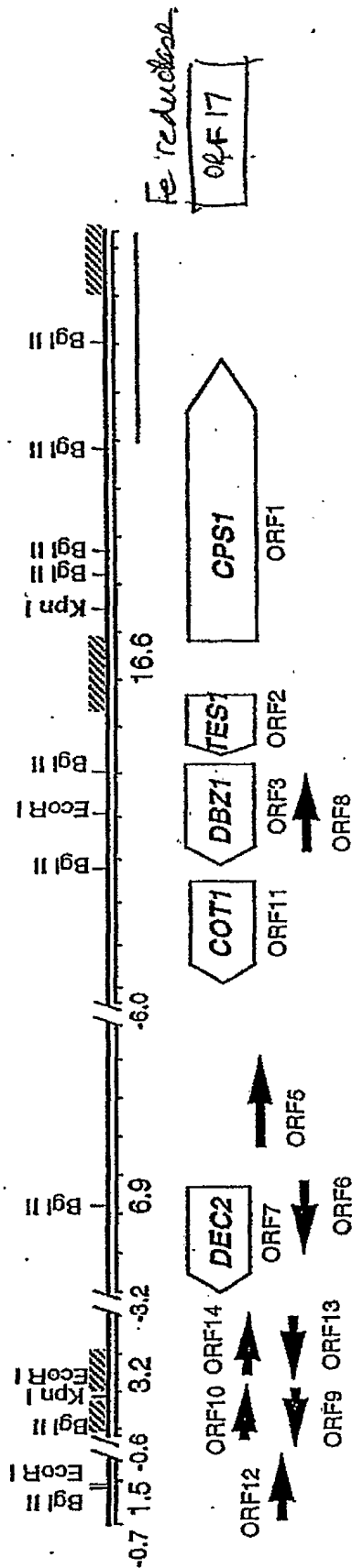
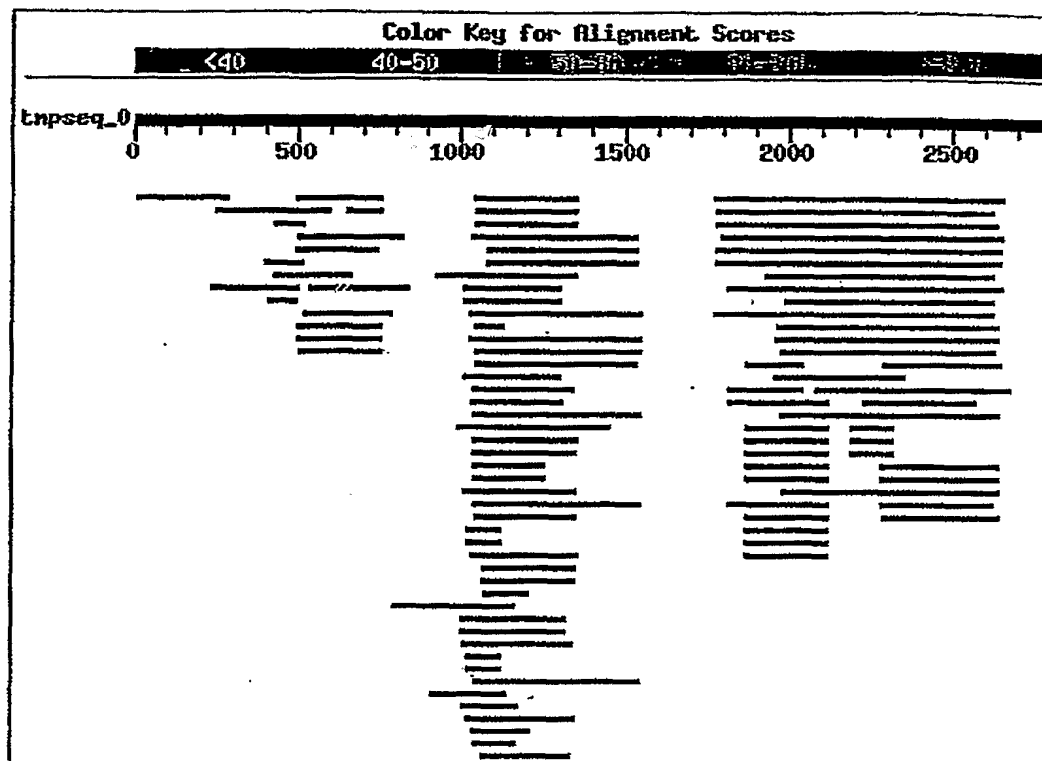


Figure 8A



QEP-15 permease, MFS transporter
QEP-16 lactase precursor

Figure 8c



Sequences producing significant alignments:

	Score	E
	(bits)	Value
ref NP_015026.1 similar to FRE2; Fre3p [Saccharomyces cere...	107	
ref NP_012702.1 [Saccharomyces cer...	106	
gb AAG09788.1 AF254143.1 (AF254143) repressed by TUP1 prote...	106	1e-21
ref NP_014458.1 Similar to ferric reductases Frelp and Fre...	100	1e-19
sp P78588 FREL CANAL PROBABLE FERRIC REDUCTASE TRANSMEMBRAN...	93	2e-17
emb CAB45649.1 (AJ387722) ferric reductase [Candida albicans]	91	6e-17
ref NP_013315.1 Ferric (and cupric) reductase; Frelp [Sacc...	91	6e-17
ref NP_015029.1 similar to FRE2; Fre5p [Saccharomyces cere...	87	7e-16
ref NP_014489.1 Fre7p [Saccharomyces cerevisiae] >gi 39136...	68	7e-10
ref NP_013049.1 similar to FRE2; Fre6p [Saccharomyces cere...	67	1e-09
sp Q04800 FRP1 SCHPO FERRIC REDUCTASE TRANSMEMBRANE COMPONE...	55	3e-06
pir T40777 ferric reductase transmembrane component - fiss...	54	7e-06
emb CAB91820.1 (AL356192) related to ferric reductase [Neu...	44	0.007
gb AAC39478.1 (AF055356) respiratory burst oxidase protein...	41	0.049
emb CAB70727.1 (AL137404) hypothetical protein [Homo sapiens]	41	0.065
pir S23737 proline-rich protein precursor - kidney bean >g...	41	0.085
gb AAF46035.1 (AE003434) CG15784 gene product (Drosophila ...	41	0.085
ref NP_040968.1 replicase protein (putative); putative [Eg...	41	0.085
pir RRWP6M genome polyprotein - eggplant mosaic virus	41	0.085
pir T01456 extensin homolog F2401.18 - Arabidopsis thalian...	40	0.11
dbj BAA95995.1 (AB040904) KIAA1471 protein [Homo sapiens]	40	0.11
pir T14756 hypothetical protein DKFZp564F0923.1 - human (f...	40	0.15
gb AAB70928.1 (AF020261) proline rich protein [Santalum al...	39	0.25
pir T04859 extensin homolog F28A21.80 - Arabidopsis thalia...	39	0.25
gb AAF87118.1 AC006434.14 (AC006434) F10A5.23 [Arabidopsis ...	39	0.25
pir S51798 gamma-kafirin precursor - sorghum	39	0.25

Figure 9A

emb CAA44347.1 	(X62480) gamma-kafirin preprotein [Sorghum ...	39	0.25
dbj BAA99921.1 	(AP001306) contains similarity to cell wall...	38	0.43
sp P13983 EXTN TOBAC	EXTENSIN PRECURSOR (CELL WALL HYDROXYP...	38	0.43
ref NP 013148.1 	Ylr047cp [Saccharomyces cerevisiae] >gi 21...	38	0.56
pir T14355	protein-tyrosine-phosphatase (EC 3.1.3.48) TD14...	38	0.74
gb AAA73078.1 	(M73688) [Sorghum bicolor endosperm tissue m...	37	0.96
sp P48038 ACRO RABIT	ACROSIN PRECURSOR >gi 1085468 pir S47...	34	1.2
pir T00264	high carbon dioxide response protein 2 - Chloro...	37	1.3
gb AAC82365.1 	(AF055904) unknown [Myxococcus xanthus]	37	1.3
pir A54523	histidine-rich protein - Plasmodium lophurae (f...	37	1.3
pir T28682	hypothetical protein - Streptomyces coelicolor ...	36	1.7
pir T06291	extensin homolog T9E8.80 - Arabidopsis thaliana...	36	1.7
sp P06599 EXTN DAUCA	EXTENSIN PRECURSOR >gi 82047 pir A243...	36	2.2
dbj BAA97321.1 	(AB020754) gene_id:MYN8.5-pir T34137-simil...	36	2.2
pir S25299	extensin precursor - tomato >gi 170444 gb AAA34...	36	2.2
gb AAF50413.1 	(AE003555) Gug gene product [Drosophila mela...	36	2.8
pir T33997	hypothetical protein W03G1.5 - Caenorhabditis e...	36	2.8
gb AAD55979.1 	AF159296 1 (AF159296) extensin-like protein [...	36	2.8
gb AAD41978.1 	AC006438 10 (AC006438) unknown protein (Arabi...	36	2.8
emb CAA76070.1 	(Y16104) overlapping protein [Physalis mott...	36	2.8
gb AAF34752.1 	AF217844 1 (AF217844) GRUNGE [Drosophila mela...	36	2.8
gb AAC39477.1 	(AF055355) respiratory burst oxidase protein...	35	3.7
ref NP 031989.1 	ecotropic viral integration site 1 [Mus mu...	35	3.7
dbj BAB08752.1 	(AB017063) respiratory burst oxidase protei...	35	3.7
gb AAF82153.1 	AC034256 17 (AC034256) Contains similarity to...	35	3.7
dbj BAA21572.1 	(AB002344) KIAA0346 [Homo sapiens]	35	3.7
sp P14404 EVII1 MOUSE	ECOTROPIC VIRUS INTEGRATION 1 SITE PRO...	35	3.7
emb CAB76093.1 	(AL157956) putative oxidoreductase [Strepto...	35	3.7
emb CAA04777.1 	(AJ001482) Evildelta 105 [Mus musculus]	35	3.7
sp P24152 EXTN SORBI	EXTENSIN PRECURSOR (PROLINE-RICH GLYCO...	35	3.7
gb AAF53661.1 	(AE003658) CG15157 gene product [Drosophila ...	35	4.9
gb AAF21492.1 	U91669 1 (U91669) merozoite surface antigen 2...	35	4.9
gb AAF51119.1 	(AE003580) CG3304 gene product [Drosophila m...	35	4.9
gb AAG21895.1 	AF237962 1 (AF237962) NADH/NADPH thyroid oxid...	35	4.9
sp P20186 YT35 STRFR	HYPOTHETICAL 35.5 KD PROTEIN IN TRANSP...	35	4.9
pir PQ0475	pistil extensin-like protein (clone pMG04) - co...	35	4.9
gb AAF73291.1 	AF155232 1 (AF155232) extensin [Pisum sativum]	35	4.9
gb AAD37405.1 	AF148222 1 (AF148222) merozoite surface prote...	35	4.9
pir A48232	cysteine-rich extensin-like protein 1 precursor...	35	4.9
ref NP 057309.1 	Rhd type IIIa protein [Homo sapiens] >gi 6...	35	4.9
gb AAC38842.1 	(AF010462) merozoite surface protein 2 [Plas...	34	6.4
pir T05530	cytochrome b245 beta chain homolog F13M23.230 -...	34	6.4
gb AAC17609.1 	(AC002131) Strong similarity to extensin-lik...	34	6.4
gb AAA16424.1 	(L13855) UL3.5 [Pseudorabies virus]	34	6.4
gb AAF64002.1 	(AF217011) merozoite surface protein 2 [Plas...	34	6.4
ref XP 001241.1 	similar to hypothetical protein FLJ20337 (...)	34	6.4
gb AAC18092.1 	(AF056965) mutant membrane protein RhCe [Hom...	34	6.4
pir T10863	extensin precursor - kidney bean >gi 727264 gb ...	34	6.4
emb CAB72300.1 	(AL031284) dJ469D22.1 (Rhesus blood group, ...)	34	6.4
pir B29356	hydroxyproline-rich glycoprotein precursor (clo...	34	6.4
dbj BAB01951.1 	(AP002048) extensin-like protein [Arabidops...	34	6.4
gb AAF34294.1 	AC005941 6 (AC005941) L5204.7 [Leishmania major]	34	6.4
gb AAD24546.2 	(AF116856) neurocan core protein precursor [...	34	6.4
pir S40517	erythrocyte membrane protein - human	34	6.4
ref NP 057208.1 	Rhesus blood group, D antigen; Rhesus syst...	34	6.4
pir S40516	erythrocyte membrane protein - human	34	6.4
emb CAB09722.1 	(Z97026) rhesus D category VI type III prot...	34	6.4
gb AAD25300.1 	AF088276 1 (AF088276) NADPH oxidase, gp91; ph...	34	8.4
gb AAB34660.1 	RhPI-2e=Rhesus blood group antigen isoform {...	34	8.4
pir T25800	C2H2-type zinc finger domain, WT1 homolog - Cae...	34	8.4

<u>pir</u> <u>S12549</u> hypothetical protein - human herpesvirus 4	<u>34</u>	8.4
<u>gb</u> <u>AAF57886.1</u> (AE003804) CG17288 gene product [Drosophila ...	<u>34</u>	8.4
<u>gb</u> <u>AAD52161.1</u> <u>AF143503.1</u> (AF143503) nitric oxide synthase [...	<u>34</u>	8.4
<u>dbj</u> <u>BAA35135.1</u> (AB008227) Extensin [Adiantum capillus-vene...	<u>34</u>	8.4
<u>gb</u> <u>AAB34659.1</u> RhPI-1d=Rhesus blood group antigen isoform {...	<u>34</u>	8.4
<u>emb</u> <u>CAB65664.1</u> (AJ252251) glycoprotein G-2 [human herpesvi...	<u>34</u>	8.4
<u>ref</u> <u>NP_044534.1</u> virion glycoprotein G [human herpesvirus 2...	<u>34</u>	8.4
<u>pir</u> <u>S78480</u> Rhesus blood group antigen-like protein isoform...	<u>34</u>	8.4
<u>ref</u> <u>NP_005057.1</u> splicing factor proline/glutamine rich (po...	<u>34</u>	8.4
<u>gb</u> <u>AAF59500.1</u> (AC024805) Hypothetical protein Y51H7C.a [Ca...	<u>34</u>	8.4
<u>pir</u> <u>I52615</u> gene RhD protein - human >gi 999310 gb AAB34852...	<u>34</u>	8.4
<u>dbj</u> <u>BAA81899.1</u> (AB018966) Rh blood group D antigen (RhD) [...	<u>34</u>	8.4
<u>ref</u> <u>NP_065231.1</u> Rhesus blood group, CcEe antigens; Rhesus ...	<u>34</u>	8.4
<u>pir</u> <u>T29299</u> hypothetical protein C50F7.2 - Caenorhabditis e...	<u>34</u>	8.4

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

	R Y W V L L C G S I L L C C L S G A S G S T A L T S D Y G K	Majority
	10 20 30	
1	R - - - - - L R S A R G T T W L H S D Y I H	Fered.pro
1	M Y W V L L C G S I L L C C L S G A S A S P A K T K M Y G K	ScFre3p.PRO
	L N L V L T D A C T G V L G G A T W E Y S S D D L Y S G S A	Majority
	40 50 60	
18	N N I - - - - N C T G P H S G A - - - - - P G S A	Fered.pro
31	L P L V L T D A C M G V L G E V T W E Y S S D D L Y S S P A	ScFre3p.PRO
	C T Y S P A A Q S M L Y C I F E S L A E A G Y S L S G L L K	Majority
	70 80 90	
34	P Q Y S P I A Q - - - - - F P P L A P A S M S L S G L L R	Fered.pro
61	C T Y E P A L Q S M L Y C I Y E S L N E K G Y S N R T F E K	ScFre3p.PRO
	S F A A I K E A C A A K T D L L S N W T A A D F Y N M L N N	Majority
	100 110 120	
58	S - - - - R E A P A A K R H L L S N W N A A Q F - - - - -	Fered.pro
91	T F A A I K E D C A Y Y T D N L Q N M T N A D F Y N M L N N	ScFre3p.PRO
	G T T Y I I Q Y S E G S A E L T Y S I G L T G V V Q V G N F	Majority
	130 140 150	
78	- - - - - E E L K Y S Y G L T G V D Q V G N F	Fered.pro
121	G T T Y I I Q Y S E G S A N L T Y P I E M D A Q V R E N Y Y	ScFre3p.PRO
	L S V D G F L Y A L I G I G H T Y G G I L L A L R V G V M V	Majority
	160 170 180	
96	L W V D T F L Y M L I G I S - - - - G M L L M L R I S N M V	Fered.pro
151	Y S Y H G F - Y A N Y D I G H T Y G G I I C A Y F V G V M I	ScFre3p.PRO
	L A S I L H S L A Y G S F K T A L F E T N R L V R Y P W V N	Majority
	190 200 210	
122	W K H S R H I T A M G S P R Q K Y W E T N R T S W W P W L N	Fered.pro
180	L A S I L H Y L S Y T P F K T A L F K - Q R L V R Y - - V R	ScFre3p.PRO

Figure 9B

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

	R H L L V A T L W G K K H A A S F S I L S A I T N G T L P G	Majority
	220 230 240	
152	R H I L V A P L W K K K H N A Q F Q I S S A I D N G T L P G	Fered.pro
207	R Y L T I P T I W G K - H A S S F S Y L K I F T - G F L P T	ScFre3p.PRO
	R S E G V I L L G Y V G L N V A F C L A L G Y Q Y D P Y D V	Majority
	250 260 270	
182	R W H T I M L L I Y V G L N V A W C L A L - - - - P Y D V	Fered.pro
235	R S E G V I I L G Y L V L H T V F - L A Y G Y Q Y D P Y N L	ScFre3p.PRO
	L F D S H R E T L A A Y V A G R S G V L A A A N L I L I A L	Majority
	280 290 300	
207	L - - D H R E T L A A - L R G R S G T L A A L N L I P T I L	Fered.pro
264	I F D S R R E Q I A R Y V A D R S G V L A F A H F P L I A L	ScFre3p.PRO
	F A G R N N F L I S L L G V S Y T S F I L F H K W A G R I T	Majority
	310 320 330	
234	F A L R N N P L I S L L Q V S Y D D F N L F H R W A A R I T	Fered.pro
294	F A G R N N F L E F I S G V K Y T S F I M F H K W L G R M M	ScFre3p.PRO
	I A D A V V H G A A Y L S N S V A G G G W A A V V A A L H T	Majority
	340 350 360	
264	I A E A I V H T A A W L Y N T K A G G G W H A V V A A L H T	Fered.pro
324	F L D A V I H G A A Y T S Y S V F Y K D W A A S - - - - K	ScFre3p.PRO
	E G S Y G W G F G G V A A L T I V G V Q A F F S L A M F R K	Majority
	370 380 390	
294	E G S Y G W G M G G T V A F T F I G I Q A W - - - S P F R H	Fered.pro
349	E E T Y - W Q F G - V A A L C I V G V M V F F S L A M F R K	ScFre3p.PRO
	A F Y E A F L N L H R V I V L G A L L G L Y T H L E H V V A	Majority
	400 410 420	
321	A F Y E T F L N I H R V M V I A A L L G L Y K H L E - L H A	Fered.pro
377	F F Y E A F L F L H - - I V L G A L F - F Y T C W E H V V E	ScFre3p.PRO

Alignment Report of Untitled, usir - Clustal method with PAM250 resid weight table.

	L S G V E W I Y A A I A I W A A D R L L R L V S V S Y F G F	Majority
	430 440 450	
350	L P Q V P W M Y L I F I F W A A E W F L R L C S I C Y Y G F	Fered.pro
404	L S G I E W I Y A A I A I W T I D R L I R I V R V S Y F G F	ScFre3p.PRO
	S L K Q R S S A S V E A V G G D A V R V T V N R P V R L W T	Majority
	460 470 480	
380	S L K Q R S S I T V E A L P G E A V R L T I N M - V R E W T	Fered.pro
434	P - - - - K A S L Q L V G D D I I R V T V K R P V R L W K	ScFre3p.PRO
	A K P G Q H V H V S F L H H L S L W S S H P F S V L D S I I	Majority
	490 500 510	
409	P R P G C H V H M - W M P R L S L W S S H P F S V - - - -	Fered.pro
459	A K P G Q Y V F V S F L H H L Y F W Q S H P F T V L D S I I	ScFre3p.PRO
	K D G E L T I I L K E K K G V T K L V K K Y V C C N G G K A	Majority
	520 530 540	
433	- -	Fered.pro
489	K D G E L T I I L K E K K G V T K L V K K Y V C C N G G K A	ScFre3p.PRO
	S M R L A I E G P Y G S S S P V N N Y D N V L L L T G G T G	Majority
	550 560 570	
433	- -	Fered.pro
519	S M R L A I E G P Y G S S S P V N N Y D N V L L L T G G T G	ScFre3p.PRO
	L P G P I A H A I K L G K T S A A T G K Q F I K L V I A V R	Majority
	580 590 600	
433	- -	Fered.pro
549	L P G P I A H A I K L G K T S A A T G K Q F I K L V I A V R	ScFre3p.PRO
55	G F N V L E A Y K P E L M C L E D L N V Q L H I Y N T M E A	Majority
	610 620 630	
433	- -	Fered.pro
579	G F N V L E A Y K P E L M C L E D L N V Q L H I Y N T M E V	ScFre3p.PRO

Alignment Report of Untitled, using Clustal method with PAM250 residue weight table.

	W A A T P D D S L E I S Q Q D E K A D G K G V V M A T T L E	Majority
	640 650 660	
434	W A A T P D R R L - - - - -	Fered.pro
609	P A L T P N D S L E I S Q Q D E K A D G K G V V M A T T L E	ScFre3p.PRO
	Q S D D A V E F G G T V H H D G R P T V E K L L Q E V G T L	Majority
	670 680 690	
443	Q R D D A S H F G R R R H H D Q W P T - - - - - Q E I K T N	Fered.pro
639	Q S P N P V E F D G T V F H H G R P N V E K L L H E V G D L	ScFre3p.PRO
	N G S L A V V C C G P P V F V T E V R A Q G A I L V L E K P	Majority
	700 710 720	
468	Q S H M P - - C - - - - P Y R T H L R A Q G D I	Fered.pro
669	N G S L A V V C C G P P V F V D E V R D Q T A N L V L E K P	ScFre3p.PRO
	A K A I E Y F E E Y Q S W	Majority
	730	
485		Fered.pro
699	A K A I E Y F E E Y Q S W	ScFre3p.PRO

Decoration 'Decoration #1': Shade (with solid black) residues that match the Consensus exactly.

Figure 9C

CCTTGGTAGTGCACTGGGCGTGTGAATAGCTGTCAGCACGCCCCGCTGGCGGTTTCGCGCCATGGTGGAGATTTTGCACGC
 GACATGGACGACGACGGCCTCGGCAGCGTGAGGAACATGTCAAAATGAACCCAGGGGTGCATCAAAGCCGTTTACCTG
~~AACAGATGAGTGGGATCTCTGCGGGATGGGTGATGAAGTTGACTCGCTTGGACGACGGTTTGGGGGCGAGGCTAGAGCC~~
 GCACATGTTCATCGGCCGGGCATGGCGTCGGGGCCTGCACAGTTCTTGCAGAGAGGGCGCGAAAGAGGGACGAGACAGGC
 GGGTGTGGCGGGCGATAGTGCGGGCGAGAGGGAAAGAGAGTTGGAGAGAGGGGAACCATGGTGTGTTGCTCGCGGCGTCGC
 GCATCGTGCAGGGCGCGTCGGCCCATCATGGCATCGTCATTGGCATCATCGGTACGTACATGGCCGCAAGCATCATGAG
 CACAGCGGCCATTTCATTTCTCGTGGCGTGGTGGCCGTGGCAAGCTGGTGTCTGTGGCGCGGCGTCTGCGTGTGTGG
 ATCGCCAGGCACCGCGGCGTGATTGGCACAGGCAGGCGAGGGCGGGCATGGATGGATGCAGCGTGTGCAGTCAATCCG
 AAGCACAGCGCAGGGCCGGCAGCCGGCACGTCTGTGGTGGGCGTGTGCAGGCTGGGCGGCGGTGCCGAGTGGGATGGGG
 ATGATAGGATGGATGGGCGAGTCTGGCGAGGGTGGGCTGAGGGGCCCTTGGTTTCATGGCGCTTCCAGGCACGCGGGCCTG
 AGTGGCCAGCATTCGAGCCTCCGCCGCCCGTTTACCCGCTAGCAGCTGGTGGACAGCCCAAGCCAATCGTCCGC
 CTGGCCCTGGCGCCCCCCTGGTCTGCTGCCGGCCAGGCCACCCCTCGCGCGAACCTCATATCCACTGTTTCCCTCGCT
 CCTCTCCGACCTCAGCTCGCGCTCCCTGCTGTCTGGCTCTCTGCCACATTCTGTTTCCACTGTTTCCCTCACTCGCTAC
 GGAACCAAAAAACAGCTCCCTGCCCGCTGCTGCTTAACTTACCCACCCCCCGCGCCCGTCAACCATTC
 CAAGCCCATTCACCCAGCCGCGCCCTGGCGAGCCATCAATCATTCAACACACAACACTCCCTCCCTCGTTTCATTACG
 CCCGATACTGTGCAGGCCACCCCGCCCTCTGACTCCATTGTGCACACAGCACTGCTGCCCGCCACTTACCATTTGTA
 ATGCATTTCTGTGACCACGCTGTCTCTCACTCCGCCAACGCCCATCGCTTCGCCTNTCGACTCCGAAGCGCCCGTG
 GGACCACCTGGCTACACAGCGACTACATTCACAACAACATCAACTGCACCGGCCCTCACTCAGGAGCACCTGGTAGCGC
 GCCTCAATACTCGCCCATCGCCCAATTCCCCCTTTAGCACCCGCCAGCATGTCTCTCTCCGGCCTGCTGCGCTCGCGG
 GAGGCACCCGCTGCCAAGCGTCACTCTCTCAACTGGAATGCCGCCAGTTTTGAGGAGCTCAAGTACTCGTACGGCC
 TCACTGGTGTGACCAAGTCGGCAACTTCTTGTGGGTGACACCTTTCTCTACATGCTCATTGGCATCTCTGGCATGCT
 CCTCATGCTCCGCATCTCCAACATGGTCTGGAAGCACAGCCGCGCATCACCGCAATGGGAAGCCCAAGGCAAAAGTAC
 TGGGAGACCAACCGAACAAGCTGGTGGCCCTGGCTCAACCGCCACATCTCTCGTCCGCCCGCTCTGGAAGAAGAAGCACA
 ACGCCAGTTCAGATCAGCAGCGGATTGACAACGGAACCTCCCTGGAAGATGGCACACCATCATGCTCTCTCATCTA
 CGTGGCCCTCAACGTTGCATGGTGCCTTGGCCCTCCCTACGAGCTCTCGACCACAGGGAGACGCTCGCCGCCCTTCGT
 GGACGCTCTGGAACCCCTCGCCGCCCTCAACCTCATCCCCACCATCTCTTCGCCCTCCGCAACAACCCCTCACTCCCT
 TCTCCAGGTCTCGTACGACGACTTCAACCTTTTCCACCGCTGGGCTGCCCGAATCACCATTGCCGAGGCCATTGTCCAC
 ACTGCCGCTTGGTGTACAAACACCAAGGCTGGCGGTGGATGGCACGCCGTCGTAGCTGCCCTCCACACCGAGGGCTCTT
 ACGGATGGGGCATGGGCGGAACCTGTCGCTTACCTTCATCGGCATCCAGGCTGGTCCCCATTCCGTCACGCCCTTTTA
 CGAGACCTTTCTCAACATCCACCGCGTCATGGTCATTGCTGCTCTCTCGGCTTGTACAAGCACCTGGAGCTGCACGCT
 CTGCCCCAGGTCCCATGGATGTACCTCATCTTCATCTTCTGGCGGCTGAGTGGTTCTCCGCTGTGCTCCATCTGCT
 ACTACGGCTTCAGCCTGAAGCAACGCTCTTCATACCGTCGAGGCTTGGCTGGCGAAGCTGTCCGTCTAACCATCAA
 CATGGTCCGCGAATGGACCCCCCGTCCCGGATGTACGTGCACATGTGGATGCCCTCGCCTCTCCCTATGGTCTTCGCAT
 CCATTTTCCGTGCGCTGGGCTGCGACCCCTGACCGACGACTCAAAGAGATGACGCTTCCCACTTTGGAAGGCGACGTC
 ACCATGATCAATGGCCAACCCAGGAAATCAAAACAAATCAGTCTCATATGCCGTGCCCGTACCGGACTCACCTTAGAGC
 ACAGGGCGATATTC

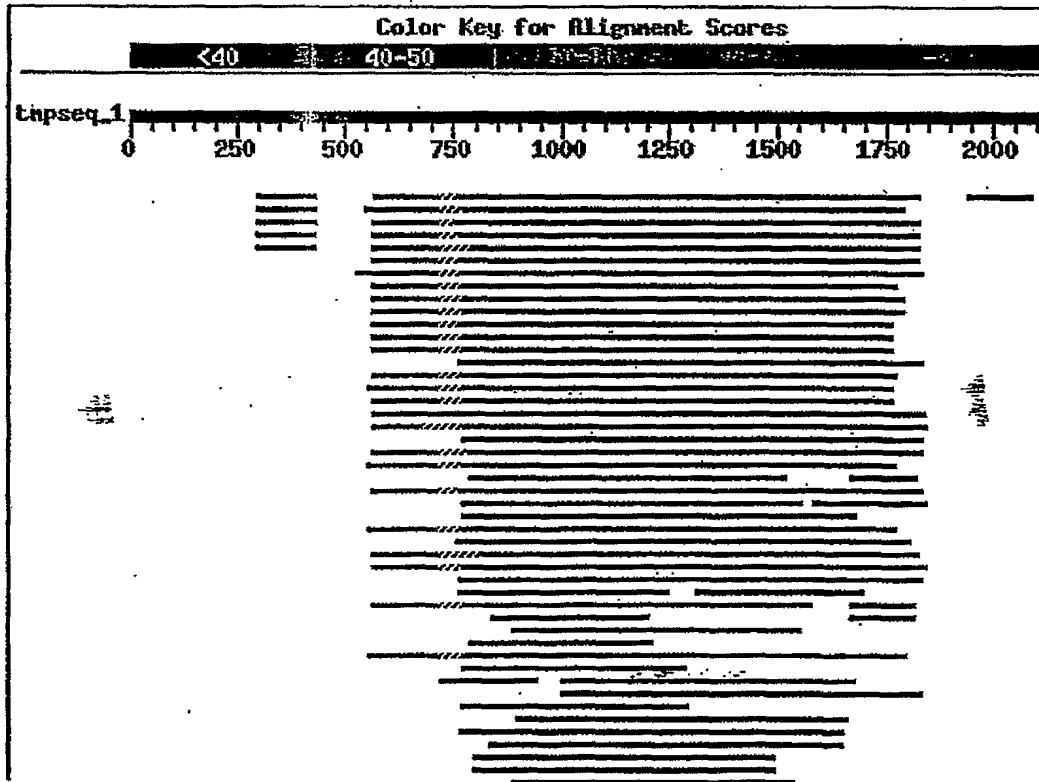
RLRSARGTTWLHSDYIHNNINCTGPHSGAPGSAPQYSPIAQFPPLAPASMSLSGLLSREAPAAKRHLLSNWNAAQFEE
 LKYSYGLTGVDQVGNFLWVDTFLYMLIGISGMLLMRLISNMVWKHSRHI TAMGSPRKQYWE TNRTSWWPWLN RHILVAP
 LWKKKHNAQFQISSAIDNGTLPGRWHTIMLLIYVGLNVAWCLALPYDVL D HRET LAALRGRSGT LAALNLIPTILFALR
 NNPLISLLQVSYDDFNLFHRWAARITIAEAI VHTAAWLYNTKAGGGWHAVVAALHTEGSYWGMMGTVAFTFIGIQAWS
 PFRHAFYETFLNIHRVMVIAALLGLYKHLELHALPQVPWMYLIFIFWAAEWFLRLCSICYYGFSLKQRSSITVEALPGE
 AVRLTINMVREWT PRPGCHVHMWMPRLSLWSSHPFSVAWAATPDRRLQRDDASHFGRRRHHDQWPTQEIKTNQSHMPCP
 YRTHLRAQGD I

Figure 9D

Figure 10

Distribution of 127 Blast Hits on the Query Sequence

Mouse-over to show define and scores. Click to show alignments



.../blast.cgi?RID=972605641-1457-10805&DESCRIPTIONS=100&ALIGNMENTS=50&ALIGNMEN10/26/00

Sequences producing significant alignments:

	Score (bits)	E Value
emb CAB99379.1 (AL390354) conserved hypothetical protein [...]	218	7e-65
emb CAB65616.1 (AL136078) probable membrane transporter [S...]	192	1e-57
emb CAB91174.1 (AL355920) putative MFS allantoate permease...	179	4e-55
sp O10097 YAOI SCHPO PUTATIVE TRANSPORTER C11D3.18C >gi 749...	188	1e-53
ref NP_011776.1 Tnalp is a high affinity nicotinic acid pl...	165	4e-49
emb CAB63540.1 (AL133521) putative transporter [Schizosacc...]	151	2e-43
qb AAG07513.1 AE004829_3 (AE004829) probable MFS transporte...	120	3e-31
qb AAG05602.1 AE004648_1 (AE004648) probable MFS transporte...	116	2e-30
pir D64995 hypothetical protein b2246 - Escherichia coli (...)	116	2e-29
sp P76470 YFAV ECOLI HYPOTHETICAL 46.3 KD PROTEIN IN GLPC-A...	116	2e-29
sp P70786 TUB3 AGRVI PUTATIVE TARTRATE TRANSPORTER >gi 9843...	116	6e-29
sp O44470 TUB4 AGRVI PUTATIVE TARTRATE TRANSPORTER >gi 8052...	114	2e-28
emb CAB61275.1 (AL132991) putative transporter protein [St...]	107	6e-27
sp O13880 YE1G SCHPO PUTATIVE TRANSPORTER C1B3.16C >gi 7493...	121	7e-27
emb CAB66219.1 (AL136503) probable transmembrane transport...	101	3e-24
emb CAB75122.1 (AL139075) transmembrane transport protein ...	97	2e-23
ref NP_012100.1 Yil166cp >gi 731893 sp P40445 YIQ6 YEAST P...	108	2e-22
emb CAA86046.1 (Z37980) hypothetical 4-hydroxyphenylacetat...	88	2e-21
pir T41604 probable membrane transport protein - fission y...	83	3e-21
pir T39680 probable allantoate permease - fission yeast (S...	91	4e-20
qb AAD53495.1 AF144422_2 (AF144422) HpaX [Salmonella dublin]	80	8e-20
ref NP_013104.1 Ylr004cp >gi 2132659 pir S64826 probable ...	90	2e-19
pir T41345 probable allantoate permease - fission yeast (S...	91	4e-17
sp O13879 YE1F SCHPO PUTATIVE TRANSPORTER C1B3.15C >gi 7493...	91	4e-17
ref NP_012686.1 allantoate permease; Dal5p >gi 118233 sp P...	74	1e-16
sp Q05181 PHT1 PSEPU PHTHALATE TRANSPORTER >gi 295708 dbj B...	87	9e-16
pir C70818 probable ABC transporter - Mycobacterium tuberc...	71	3e-15
qb AAD41517.1 AF152094_1 (AF152094) phthalate transporter [...]	83	1e-14
ref NP_009957.1 Amino acid permease; Fen2p >gi 140479 sp P...	81	4e-14
ref NP_009333.1 putative permease; Seolp >gi 731298 sp P39...	78	5e-13
qb AAG05650.1 AE004652_3 (AE004652) probable 2-ketogluconat...	67	5e-13
pir T40485 transmembrane transporter liz1p - fission yeast...	69	2e-10
pir T40140 transmembrane transporter liz1p - fission yeast...	68	4e-10
ref NP_014479.1 Yoll163wp >gi 2132861 pir S66862 probable ...	63	1e-08
qb AAD03552.1 (AF095748) putative phthalate permease C-te...	61	4e-08
ref NP_013045.1 Yll055wp >gi 1077324 pir S50965 probable ...	52	9e-08
pir A71619 membrane transporter PFB0275w - malaria parasit...	55	3e-06
dbj BAA87280.1 (AB027976) Hypothetical protein [Schizosacc...]	51	4e-05
pir T12997 hypothetical protein T21L8.170 - Arabidopsis th...	50	7e-05
dbj BAA13885.1 (D89224) similar to Saccharomyces cerevisia...	50	1e-04
qb AAD46810.1 AF157643_4 (AF157643) putative transporter pr...	49	2e-04
ref NP_005486.1 Na/PO4 cotransporter >gi 4587207 dbj BAA76...	48	4e-04
dbj BAA95074.1 (AB041591) unnamed protein product [Mus mus...]	47	6e-04
sp P39398 YJLJ ECOLI HYPOTHETICAL 49.4 KD PROTEIN IN TSR-MD...	47	6e-04
qb AAD49570.1 AF135037_1 (AF135037) nitrate transporter [Cy...]	47	8e-04
qb AAF55770.1 (AE003730) CG4288 gene product [alt 1] [Dros...]	46	0.002
qb AAD49571.1 (AF135038) nitrate transporter [Cylindrothec...]	46	0.002
qb AAD49572.1 AF135039_1 (AF135039) nitrate transporter [Cy...]	45	0.002
qb AAC13878.1 (U39735) high molecular weight basic nuclear...	44	0.005
sp P31457 DGOT ECOLI D-GALACTONATE TRANSPORTER	43	0.016
pir D65171 hypothetical 48.8 kD protein in ibpA-gyrB inter...	43	0.016
pir T15201 hypothetical protein F12B6.2 - Caenorhabditis e...	42	0.028
qb AAF97770.1 AF244578_1 (AF244578) membrane glycoprotein S...	41	0.062

ref NP_011579.1	H ⁺ -biotin symporter; Vhtlp >gi 1723674 sp ...	40	0.081
sp Q9Z7N9 UHPT_CHLPN	PROBABLE HEXOSE PHOSPHATE TRANSPORT PR...	39	0.14
gb AAD25166.1 AF030397_4	(AF030397) putative 4-chlorobenzoa...	39	0.14
emb CAB88550.1	(AL353819) related to carboxylic acid trans...	39	0.18
gb AAF78822.1	(AF042490) 4-chlorobenzoate transport protei...	39	0.24
gb AAG03624.1 AE004461_7	(AE004461) 4-hydroxybenzoate trans...	38	0.31
pir S14183	DNA-directed RNA polymerase (EC 2.7.7.6) larges...	38	0.54
gb AAB59301.1	(L20696) meiotin-1 [Lilium longiflorum]	38	0.54
pir S14182	DNA-directed RNA polymerase (EC 2.7.7.6) larges...	37	0.71
pir S14181	DNA-directed RNA polymerase (EC 2.7.7.6) larges...	37	0.71
ref NP_037488.1	monocarboxylate transporter 3 >gi 6093322 ...	37	0.71
pir T07796	DNA-directed RNA polymerase (EC 2.7.7.6) larges...	37	0.71
emb CAA36734.1	(X52493) DNA-directed RNA polymerase [Glyci...	37	0.71
gb AAF59171.2	(AE003839) CG8791 gene product [Drosophila m...	37	0.71
gb AAF08182.1	(AF130226) NADH dehydrogenase subunit F [Que...	37	0.93
sp Q28722 NPT1_RABIT	RENAL SODIUM-DEPENDENT PHOSPHATE TRANS...	36	1.2
pir F75580	probable sugar transporter - Deinococcus radiod...	36	1.2
pir S56583	hypothetical protein f261b - Escherichia coli >...	36	1.6
sp Q01636 LMBV_CHICK	LAMININ BETA-1 CHAIN VARIANT (LAMININ ...	36	2.1
gb AAF49470.2	(AE003527) CG4877 gene product [Drosophila m...	35	2.7
gb AAD00528.1	(U82223) titin [Sus scrofa]	35	2.7
gb AAF53319.1	(AE003640) BG:DS07660.1 gene product [Drosop...	35	2.7
gb AAF44801.1 AE003406_6	(AE003416) hypothetical protein [D...	35	2.7
gb AAF67524.1 AF204396_1	(AF204396) monocarboxylate transpo...	35	2.7
sp Q62795 NPT1_RAT	RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPO...	35	3.6
gb AAF97769.1	(AF244577) membrane glycoprotein HP59 [Homo ...	35	3.6
ref NP_036566.1	solute carrier family 17 (anion/sugar tran...	35	3.6
gb AAC15775.1	(AF061335) oxytetracycline exporter [Strepto...	34	4.7
pir JE0378	DNA (cytosine-5-)-methyltransferase (EC 2.1.1.3...	34	4.7
pir F69443	octaprenyl-diphosphate synthase (ispB) homolog ...	34	4.7
gb AAD04032.1	(AF079900) tetracycline efflux protein [Stre...	34	4.7
emb CAB76602.1	(AJ271264) glycerol kinase [Staphylococcus ...	34	4.7
sp F77416 HYFD_ECOLI	HYDROGENASE-4 COMPONENT D >gi 7466377 ...	34	6.1
gb AAF63452.1 AF218267_9	(AF218267) benzoate transport prot...	34	6.1
gb AAB28462.1	extensin-nodule-specific proline-rich protei...	34	6.1
emb CAB67228.1	(AJ271079) NdhF' protein [Oenothera elata s...	34	6.1
emb CAB67217.1	(AJ271079) NADH-plastoquinone oxidoreductas...	34	6.1
emb CAB76600.1	(AJ271262) glycerol kinase [Staphylococcus ...	34	6.1
emb CAB76601.1	(AJ271263) glycerol kinase [Staphylococcus ...	34	6.1
emb CAB76609.1	(AJ271271) glycerol kinase [Staphylococcus ...	34	6.1
ref NP_009852.1	Probable multidrug resistance protein; Ybr...	34	8.0
pir T27092	hypothetical protein Y51B9A.6 - Caenorhabditis ...	34	8.0
sp Q51955 PCAK_PSEPU	4-HYDROXYBENZOATE TRANSPORTER >gi 1147...	34	8.0
pir T34995	probable integral membrane efflux protein - Str...	34	8.0
gb AAC69842.1	(AF076683) unknown [Staphylococcus aureus]	34	8.0
emb CAB76606.1	(AJ271268) glycerol kinase [Staphylococcus ...	34	8.0
emb CAB76608.1	(AJ271270) glycerol kinase [Staphylococcus ...	34	8.0

SEQUENCE LISTING

<110> Syngenta Participations AG
5 Cornell Research Foundation, Inc.
 Yoder, Olen
 Turgeon, Barbara G.
 Lu, Shen-wen

10<120> Fungal Iron Reductase Gene

<130> 1360.017WO1

<150> US 60/252,732
15<151> 2000-11-22

<150> US 60/252,649
<151> 2000-11-22

20<160> 210

<170> FastSEQ for Windows Version 4.0

25<210> 1
 <211> 9
 <212> PRT
 <213> Artificial Sequence

30<220>
 <223> Motif

 <221> SITE
 <222> 3
35<223> Xaa = Glu or Asp

 <221> SITE
 <222> 4, 6, 8
 <223> Xaa = any amino acid

40

<400> 1

Val Leu Xaa Xaa Gly Xaa Gly Xaa Gly

1

5

<210> 2

5<211> 6550

<212> DNA

<213> Cochliobolus heterostrophus

<400> 2

10	tgcctgcgcc	tgtgcttgtg	cctgtggaat	gtcgcggccc	gctgctgcat	agcctatctg	60
	tacatacaac	accatcccat	cccgtttcac	ctgccttgcc	tccctcctcg	tgccacacat	120
	ccgcgcgcca	caacaccatg	gctgcgacca	accccgagct	gcaggccaaa	ctgcaggagc	180
	tggaccacga	gctcgaggag	ggcgatatta	cacaaaaagg	gtccgtactg	ctgcaccacc	240
	accgccatcc	gcctctctgc	gtgcgctaata	cagtcgcata	gctatgaaaa	acgtcgcacc	300
15	gtgctgctgt	cgcagtatct	agggcctgac	tttgctgccc	agttgcaggc	cgacctgaac	360
	cagcagaacc	cacccaacc	atccagttag	ggctctcgct	cccgcaccgc	atcctttgct	420
	attccgtccg	gtccgagtc	atcacggcga	ccacaacccc	cacatatcca	gtccccccgc	480
	cccgactcat	accatgacgc	ttccgcacag	ggccaattgg	gcgcacccat	gccatattgcg	540
	aacgcctccg	ccgctgcctc	ggggggctcg	cagtacatgg	catacccgcc	cagccaagtc	600
20	ggcgcgtttt	aagagaagca	gctgggcctg	cgtacaaatt	cgctccagcg	caattcctca	660
	cagctgtcgc	aaggaagcga	gacgtttcatt	ccacggcctc	aaacgcctga	atacaaccac	720
	tcgcgcgagc	ccaccatgat	gggcaactac	gccttcaatc	cagacaatca	gcaaagttat	780
	gatggccaat	ttggctctcc	gggagaggcc	agtcgaagga	gcaccatgct	cgaggtaaac	840
	cagggttatt	tttccgactt	cacaggccaag	cagatgcaag	acaatcgcg	ctcgtatggg	900
25	ggacccaacc	gctactcgtc	gggagatgcc	ttttctccta	ccgcgcgcat	tccacctccc	960
	atgatgaacc	ccaacgatct	ccccttgggc	gctgctgaaa	ccatgatgcc	gctagagccc	1020
	cgcgatctgc	cttttgacgt	ttacgacctt	cacaacccca	atgtcaaaat	gtcaaagttt	1080
	gacaacattg	gcgctgtctt	gcgtcaccga	agtcgcacac	agccaaggac	gactgccttc	1140
	tgggtccttg	acgcaaaagg	caaagagacg	gcgtccatca	cctgggaaaa	ggtggctagt	1200
30	cgcgcgga	aggtggccaa	agtgattcgg	gacaagagca	acctctatcg	aggcgaccgt	1260
	gtggcattag	tgtacaggga	tacagaaatc	attgattttg	tcgtggcggt	gatgggctgc	1320
	ttcattgcgg	gcgttgtagc	ggtacccatc	aatagcgctg	acgactacca	gaaactcatt	1380
	cttctcctaa	cgacaactca	agctcatctc	gcattgacca	cagacaacaa	tctcaaggcc	1440
	tttcatcgtg	acattagtca	gaaccgtctg	aaatggccga	gtggggtaga	gtggtggaag	1500
35	acgaacgagt	ttggcagcca	ccacccaag	aaacatgacg	atactccagc	tttgcaagta	1560
	ccagagggtg	cctatattga	gttctcgcgt	gcacctactg	gtgaccttcg	cggtgtgggtg	1620
	cttagtcacc	ggactattat	gcaccaaattg	gcctgcatca	gtgccatgat	tagcacgata	1680
	cccaccaacg	ctcagagcca	agacacgttc	agcactagcc	tacgggatgc	agagggaaag	1740
	ttcgttgctc	cagcaccgtc	cagaaacccc	acagaagtga	tcctcacgta	cctcgacccg	1800
40	cgcgaaagcg	ctggtctcat	tctcagtgtc	ttgtttgcag	tttatggagg	ccacaccacc	1860

	gtatggctcg	agacagcgac	catggaaacc	ccgggtctat	atgcacatct	catcaccaaa	1920
	tacaagtcca	acatactgct	agcggattac	ccaggcctca	agcgcgctgc	atacaactac	1980
	caacaggatc	caatggctac	aagaaacttc	aagaaaaaca	cagaacccaa	cttcgcctcc	2040
	gtgaagatct	gtctgattga	cacgcttacc	gtcgaactgtg	aattttcacga	aattctcggga	2100
	5gatcgatatt	tcaggccact	gcgaaaccct	agagcgcgag	aactgatcgc	gccaatgctc	2160
	tgcttgccag	aacatgggtg	aatgataata	tctgtacgcg	actggctagg	tgagagaggag	2220
	cgcattgggct	gcccgcctaag	catagcagta	gaagagtcag	ataatgatga	agatgataca	2280
	gaggataagt	atgcagcggc	aaatggctac	tccagtctta	ttggtgggtg	cactacaaag	2340
	aacaaaaagg	agaagaagaa	gaaaggcccg	acagagctta	cagaaatctt	gctggacaag	2400
10	gaagctctga	agatgaacga	agtcattgtt	ctggccattg	gagaagaagc	aagcaagcgg	2460
	gcaaacgagc	ccggcaccat	gcgagtcggt	gcctttggat	accccatacc	ggatgcgaca	2520
	ctagctattg	tagaccctga	gacaagtctt	ctatgttcac	catactcgat	aggcgagatc	2580
	tgggtagatt	cgccttcact	ctctgggtggc	ttctggcagc	tgcagaagca	tacagagacc	2640
	atthttccatg	ctcgaccata	ccgtttcgtt	gagggtagcc	ctacgccaca	gttgcttgaa	2700
15	ctcgagtttc	tgcgtactgg	actcctcggc	tttgtttag	agggaataat	atthgtcctt	2760
	ggactgtacg	aagatcgcat	cagacagcgt	gttgaatggg	tagaaaatgg	tcagcttgaa	2820
	gccgagcatc	gatacttttt	tgtgcagcac	ctggtcacaa	gcattatgaa	ggccgtgcca	2880
	aaaattttacg	actggtaagt	gagctgccaa	cagagcaagg	actgtctaac	gtgtcatagc	2940
	togtctgtttg	attcttatgt	aaatggtgaa	tacctgccaa	tcattctcat	cgagacgcag	3000
20	gccgcgcatcg	ctgcgcccac	aaaccaggt	ggaccaccac	aacaattgga	tataccattt	3060
	ttggattcac	tatctgagag	gtgcatggag	gtcctttacc	aagagcatca	tttacgggta	3120
	tactgcgtga	tgattacagc	acctaataca	cttccacgag	tcatacaagaa	cggacggcga	3180
	gaaattggca	atatgctgtg	taggagagag	tttgacaatg	gctctctgcc	ctgtgtacac	3240
	gtaaagttht	gcattgagcg	atcagtgcag	aacattgcgc	tcggtgacga	ttccgctggc	3300
25	ggcatgtggt	catttgaggc	atcaatggca	cgtcagcaat	tcttgatgct	ccaagacaag	3360
	caatactctg	gtgtcgatca	tcgcgaagtc	gtcattgacg	acaggacatc	gactccactc	3420
	aatcagttct	cgaatatcca	cgacctgatg	caatggcggtg	tatctcggca	ggccgaggaa	3480
	cttgcttact	gcaactgtcg	cggtcgagga	aaagagggca	aaggcgtcaa	ttggaagaag	3540
	tttgatcaaa	agggtgcggg	cgtagcaatg	tacctcaaga	acaagggtcaa	gggtccaggcc	3600
30	ggcgatcatc	tccttctgat	gtacacgcac	tcagaagaat	ttgttttatgc	tggtcatgca	3660
	tgthttgtgc	ttggagctgt	ttgcatacca	atggcgccaa	ttgatcagaa	ccggttgaat	3720
	gaggatgcgc	cggccttgct	gcatactctt	gcagatttca	agggtcaaagc	cattcttgctc	3780
	aacgctgacg	ttgacctctt	gatgaagatc	aagcaagtat	cgcagcacat	caaacaatcg	3840
	gccgctatcc	tcaagatcag	tgtgccaaac	acatacagca	caacaaagcc	gccaaagcaa	3900
35	tcagtggtg	gccgcgacct	caagcttaca	attcgaccgg	catggattca	ggcgggtttc	3960
	ccagtgctag	tctggacata	ctggacgccc	gatcaacgtc	gtatcgcagt	tcagctgggc	4020
	catagccaaa	tcattggcact	gtgcaaggtc	caaaaagaaa	catgccaaat	gacaagtaca	4080
	cgaccagtc	ttggttggtg	ccggagcacg	ataggacttg	gtttccttca	cacttgctctc	4140
	atgggaatct	tccttgccgc	accacatac	ctggtgtcac	ctgttgactt	tgcacaaaac	4200
40	cctaataattc	tgthccaaac	gctthtcgcg	tacaagatca	aggatgcata	tgcaacgagt	4260

caa	atg	ttg	g	acc	acg	ccat	cgc	acg	cgg	g	ctg	gga	aga	gtat	ggctct	gcac	gag	ctg	4320		
aag	aat	ctc	ca	tgat	tcg	ac	tgat	gga	aga	ccac	gcg	gtt	atg	ttt	gtaa	gtga	acatt	t	4380		
gtat	gag	agg	a	act	ttc	atga	ttg	cta	actc	aat	gcag	acc	aa	agag	tgcg	tgt	gcact	ttt	4440		
gcg	ccag	cca	a	act	tag	accc	aac	cgca	atc	aac	act	gtct	act	cac	atgt	att	gaac	cca	4500		
5at	ggtag	cat	cac	gat	cata	cat	gtgt	tatt	gag	ccag	tcg	agct	ccat	ct	cgat	gtg	cat		4560		
gct	ctgcg	ac	gcgg	cct	cgt	cat	gcc	cggt	gacc	ctgaca	cag	agcc	caa	cgct	ttg	ctc			4620		
gtcc	aag	act	cgg	gcat	gg	gcc	agt	gag	c	gcg	aaat	at	ccatt	gtcaa	ccc	agag	acc		4680		
aac	caact	gt	gct	tga	acg	cgag	tac	ggc	gag	atct	ggg	tgc	agt	ccga	ggc	gaat	gct		4740		
tata	gctt	ct	acat	gtc	gaa	agag	cgct	t	gat	gcaga	ac	gctt	caat	gg	gagg	ac	gatt		4800		
10gac	ggag	acc	caa	atgt	gcg	atat	gtt	cgt	acag	gcg	att	tagg	att	ttt	gcac	agc	gtg		4860		
acac	ggcc	caa	ttg	gac	caa	cggt	gcac	ct	gtgat	atgc	aggt	gct	ttt	ctg	gtt	gga			4920		
agc	atag	gtg	ac	act	ttt	tga	agt	caac	gga	ctga	acc	att	tct	ctat	gga	catt	gag	cag	4980		
tct	gtt	gaac	gtt	gtc	acc	g	gaat	att	gtc	cct	ggag	gct	gg	tac	gtt	tc	gatt	cgc	5040		
tgt	tatt	ttag	t	aaata	act	tta	cta	acact	ct	acag	tgt	gt	ttt	ccag	gca	ggt	ggg	ctt	5100		
15tt	gtt	gtc	gt	tgt	gga	aatc	ttcc	gac	gca	act	tcct	cgc	aag	cat	gg	t	cgt	gatt	5160		
tca	atg	caat	ttt	gaac	gag	cat	cag	ctg	gc	tatt	gac	at	tgt	ctc	gtt	gtg	caaaa	agg	5220		
gcg	act	ttcca	ccg	gtc	tcgt	ctg	ggg	cgaga	ag	caac	gcg	gg	aa	agatt	ctt	gcag	gat	ggg	5280		
tcac	acg	gaa	gat	gcg	caca	atag	ccc	agt	acag	tata	cgc	ggat	ccta	aat	gg	acag	gatt		5340		
ccc	agat	gat	cac	gga	agag	cct	ggt	ccac	ggg	ctag	cat	gact	gga	agt	atg	cct	ggg	c	5400		
20gaa	tggg	cgg	ccc	agc	cag	t	atca	agg	ccg	gg	tcg	aca	ag	gcac	cgag	t	aat	ggg	ca	5460	
tgac	agc	gac	tat	gaata	aat	ctat	ccct	tta	cac	agc	agca	acag	cag	caa	tacca	acag	c		5520		
cgg	gtat	gta	tgct	caac	ag	caagg	cat	gc	acccc	agca	aca	acac	caa	tttag	cat	gt			5580		
cca	acac	gcc	acc	aca	agg	t	ccac	ccc	aa	g	cg	taga	act	acat	gat	cct	agc	gac	gca	5640	
cacca	acaga	caacc	ggc	ac	tct	ttc	ctt	g	ccg	accc	gcg	tat	gcaga	aac	cagg	gcca	aaa		5700		
25tga	acg	agac	ggg	gc	cctac	gaac	ccat	gta	act	atca	aaaa	cgc	gtat	cat	ccg	cat	caac		5760		
aaca	ata	cga	atct	ga	agac	ggg	ggg	gag	ca	gact	cag	cg	cccc	gtg	cca	gac	gtg	ctgc	5820		
ggc	ggg	tcc	ttcat	ccgg	g	tcc	atag	agc	ag	cac	gacca	ag	cta	aca	aac	gaca	aca	ata	5880		
tgt	gga	ataa	tcg	c	gag	tac	tat	ggta	aca	gccc	atc	gt	gcag	gcg	ga	tac	acg	caag	5940		
atg	gca	atat	ccac	gag	cag	caaca	acac	g	atg	agt	tac	ac	gag	ta	atg	cgc	t	cat	atgg	6000	
30gaa	atca	agg	agc	agg	cgg	ga	ggc	agc	ggc	g	gg	tgg	cg	g	tct	ccg	agt	t	gcaa	atcgt	6060
acag	ctcc	ga	cag	c	agg	gt	gcag	atg	acg	ctt	gg	ag	acgt	gat	gcc	ctt	gct	caga		6120	
tca	at	ttt	gtc	ggg	cg	cgt	gct	gcct	ccg	ctg	gg	agc	acct	gct	gct	ggt	gct	tct		6180	
ctt	gc	agcc	ggg	ccat	gcg	cag	tag	acgg	gat	atg	cgtg	ag	ttt	ttt	ttt	t	aaat	ttc	gt	6240	
acata	gag	ac	g	ttg	tata	c	gcag	gtt	tca	aatt	aga	ga	ga	aat	atg	c	at	atcag	ctg	6300	
35tt	gtt	tcaat	gt	ctag	ttt	g	gaagg	gt	taa	cccc	cccc	tt	cccc	ctt	cc	aag	act	ttt	ct	6360	
act	tgt	ttt	gt	gtg	tatt	ta	aat	ctg	gaga	ttt	caa	atct	acat	ctc	gct	ata	cat	aggt		6420	
gtt	gtt	tgat	aac	gtag	ggg	g	caga	agg	gt	atct	cgt	gat	att	ag	act	gg	gag	ttg	catg	6480	
aat	caagg	tg	ttg	ag	caaaa	aa	agag	agag	cgg	tga	agg	g	ggg	ggg	ggg	gat	agg	tgt	gtg	6540	
cac	gtg	gctg																		6550	

<210> 3

<211> 1743

<212> PRT

5<213> *Cochliobolus heterostrophus*

<400> 3

```

Met Leu Glu Val Asn Gln Gly Tyr Phe Ser Asp Phe Thr Gly Gln Gln
  1              5              10              15
10Met Gln Asp Asn Arg Asp Ser Tyr Gly Gly Pro Asn Arg Tyr Ser Ser
      20              25              30
    Gly Asp Ala Phe Ser Pro Thr Ala Ala Ile Pro Pro Pro Met Met Asn
      35              40              45
    Pro Asn Asp Leu Pro Leu Gly Ala Ala Glu Thr Met Met Pro Leu Glu
15    50              55              60
    Pro Arg Asp Leu Pro Phe Asp Val Tyr Asp Pro His Asn Pro Asn Val
    65              70              75              80
    Lys Met Ser Lys Phe Asp Asn Ile Gly Ala Val Leu Arg His Arg Ser
      85              90              95
20Arg Thr Gln Pro Arg Thr Thr Ala Phe Trp Val Leu Asp Ala Lys Gly
      100              105              110
    Lys Glu Thr Ala Ser Ile Thr Trp Glu Lys Val Ala Ser Arg Ala Glu
      115              120              125
    Lys Val Ala Lys Val Ile Arg Asp Lys Ser Asn Leu Tyr Arg Gly Asp
25    130              135              140
    Arg Val Ala Leu Val Tyr Arg Asp Thr Glu Ile Ile Asp Phe Val Val
    145              150              155              160
    Ala Leu Met Gly Cys Phe Ile Ala Gly Val Val Ala Val Pro Ile Asn
      165              170              175
30Ser Val Asp Asp Tyr Gln Lys Leu Ile Leu Leu Leu Thr Thr Thr Gln
      180              185              190
    Ala His Leu Ala Leu Thr Thr Asp Asn Asn Leu Lys Ala Phe His Arg
      195              200              205
    Asp Ile Ser Gln Asn Arg Leu Lys Trp Pro Ser Gly Val Glu Trp Trp
35    210              215              220
    Lys Thr Asn Glu Phe Gly Ser His His Pro Lys Lys His Asp Asp Thr
    225              230              235              240
    Pro Ala Leu Gln Val Pro Glu Val Ala Tyr Ile Glu Phe Ser Arg Ala
      245              250              255
40Pro Thr Gly Asp Leu Arg Gly Val Val Leu Ser His Arg Thr Ile Met

```

	260		265		270
	His Gln Met Ala Cys Ile Ser Ala Met Ile Ser Thr Ile Pro Thr Asn				
	275		280		285
	Ala Gln Ser Gln Asp Thr Phe Ser Thr Ser Leu Arg Asp Ala Glu Gly				
5	290		295		300
	Lys Phe Val Ala Pro Ala Pro Ser Arg Asn Pro Thr Glu Val Ile Leu				
	305		310		315
	Thr Tyr Leu Asp Pro Arg Glu Ser Ala Gly Leu Ile Leu Ser Val Leu				
		325		330	
10	Phe Ala Val Tyr Gly Gly His Thr Thr Val Trp Leu Glu Thr Ala Thr				
		340		345	
	Met Glu Thr Pro Gly Leu Tyr Ala His Leu Ile Thr Lys Tyr Lys Ser				
		355		360	
	Asn Ile Leu Leu Ala Asp Tyr Pro Gly Leu Lys Arg Ala Ala Tyr Asn				
15	370		375		380
	Tyr Gln Gln Asp Pro Met Ala Thr Arg Asn Phe Lys Lys Asn Thr Glu				
		385		390	
	Pro Asn Phe Ala Ser Val Lys Ile Cys Leu Ile Asp Thr Leu Thr Val				
		405		410	
20	Asp Cys Glu Phe His Glu Ile Leu Gly Asp Arg Tyr Phe Arg Pro Leu				
		420		425	
	Arg Asn Pro Arg Ala Arg Glu Leu Ile Ala Pro Met Leu Cys Leu Pro				
		435		440	
	Glu His Gly Gly Met Ile Ile Ser Val Arg Asp Trp Leu Gly Gly Glu				
25	450		455		460
	Glu Arg Met Gly Cys Pro Leu Ser Ile Ala Val Glu Glu Ser Asp Asn				
		465		470	
	Asp Glu Asp Asp Thr Glu Asp Lys Tyr Ala Ala Ala Asn Gly Tyr Ser				
		485		490	
30	Ser Leu Ile Gly Gly Gly Thr Thr Lys Asn Lys Lys Glu Lys Lys Lys				
		500		505	
	Lys Gly Pro Thr Glu Leu Thr Glu Ile Leu Leu Asp Lys Glu Ala Leu				
		515		520	
	Lys Met Asn Glu Val Ile Val Leu Ala Ile Gly Glu Glu Ala Ser Lys				
35	530		535		540
	Arg Ala Asn Glu Pro Gly Thr Met Arg Val Gly Ala Phe Gly Tyr Pro				
		545		550	
	Ile Pro Asp Ala Thr Leu Ala Ile Val Asp Pro Glu Thr Ser Leu Leu				
		565		570	
40	Cys Ser Pro Tyr Ser Ile Gly Glu Ile Trp Val Asp Ser Pro Ser Leu				

	580		585		590
	Ser Gly Gly Phe Trp Gln Leu Gln Lys His Thr Glu Thr Ile Phe His				
	595		600		605
	Ala Arg Pro Tyr Arg Phe Val Glu Gly Ser Pro Thr Pro Gln Leu Leu				
5	610		615		620
	Glu Leu Glu Phe Leu Arg Thr Gly Leu Leu Gly Phe Val Val Glu Gly				
	625		630		635
	Lys Ile Phe Val Leu Gly Leu Tyr Glu Asp Arg Ile Arg Gln Arg Val				
	645		650		655
10	Glu Trp Val Glu Asn Gly Gln Leu Glu Ala Glu His Arg Tyr Phe Phe				
	660		665		670
	Val Gln His Leu Val Thr Ser Ile Met Lys Ala Val Pro Lys Ile Tyr				
	675		680		685
	Asp Cys Ser Ser Phe Asp Ser Tyr Val Asn Gly Glu Tyr Leu Pro Ile				
15	690		695		700
	Ile Leu Ile Glu Thr Gln Ala Ala Ser Thr Ala Pro Thr Asn Pro Gly				
	705		710		715
	Gly Pro Pro Gln Gln Leu Asp Ile Pro Phe Leu Asp Ser Leu Ser Glu				
	725		730		735
20	Arg Cys Met Glu Val Leu Tyr Gln Glu His His Leu Arg Val Tyr Cys				
	740		745		750
	Val Met Ile Thr Ala Pro Asn Thr Leu Pro Arg Val Ile Lys Asn Gly				
	755		760		765
	Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe Asp Asn Gly				
25	770		775		780
	Ser Leu Pro Cys Val His Val Lys Phe Gly Ile Glu Arg Ser Val Gln				
	785		790		795
	Asn Ile Ala Leu Gly Asp Asp Pro Ala Gly Gly Met Trp Ser Phe Glu				
	805		810		815
30	Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys Gln Tyr				
	820		825		830
	Ser Gly Val Asp His Arg Glu Val Val Ile Asp Asp Arg Thr Ser Thr				
	835		840		845
	Pro Leu Asn Gln Phe Ser Asn Ile His Asp Leu Met Gln Trp Arg Val				
35	850		855		860
	Ser Arg Gln Ala Glu Glu Leu Ala Tyr Cys Thr Val Asp Gly Arg Gly				
	865		870		875
	Lys Glu Gly Lys Gly Val Asn Trp Lys Lys Phe Asp Gln Lys Val Ala				
	885		890		895
40	Gly Val Ala Met Tyr Leu Lys Asn Lys Val Lys Val Gln Ala Gly Asp				

	900		905		910
	His Leu Leu Leu Met Tyr Thr	His Ser Glu Glu Phe Val Tyr Ala Val			
	915	920	925		
	His Ala Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala Pro Ile				
5	930	935	940		
	Asp Gln Asn Arg Leu Asn Glu Asp Ala Pro Ala Leu Leu His Ile Leu				
	945	950	955	960	
	Ala Asp Phe Lys Val Lys Ala Ile Leu Val Asn Ala Asp Val Asp His				
	965	970	975		
10	Leu Met Lys Ile Lys Gln Val Ser Gln His Ile Lys Gln Ser Ala Ala				
	980	985	990		
	Ile Leu Lys Ile Ser Val Pro Asn Thr Tyr Ser Thr Thr Lys Pro Pro				
	995	1000	1005		
	Lys Gln Ser Ser Gly Cys Arg Asp Leu Lys Leu Thr Ile Arg Pro Ala				
15	1010	1015	1020		
	Trp Ile Gln Ala Gly Phe Pro Val Leu Val Trp Thr Tyr Trp Thr Pro				
	1025	1030	1035	1040	
	Asp Gln Arg Arg Ile Ala Val Gln Leu Gly His Ser Gln Ile Met Ala				
	1045	1050	1055		
20	Leu Cys Lys Val Gln Lys Glu Thr Cys Gln Met Thr Ser Thr Arg Pro				
	1060	1065	1070		
	Val Leu Gly Cys Val Arg Ser Thr Ile Gly Leu Gly Phe Leu His Thr				
	1075	1080	1085		
	Cys Leu Met Gly Ile Phe Leu Ala Ala Pro Thr Tyr Leu Val Ser Pro				
25	1090	1095	1100		
	Val Asp Phe Ala Gln Asn Pro Asn Ile Leu Phe Gln Thr Leu Ser Arg				
	1105	1110	1115	1120	
	Tyr Lys Ile Lys Asp Ala Tyr Ala Thr Ser Gln Met Leu Asp His Ala				
	1125	1130	1135		
30	Ile Ala Arg Gly Ala Gly Lys Ser Met Ala Leu His Glu Leu Lys Asn				
	1140	1145	1150		
	Leu Met Ile Ala Thr Asp Gly Arg Pro Arg Val Asp Val Tyr Gln Arg				
	1155	1160	1165		
	Val Arg Val His Phe Ala Pro Ala Asn Leu Asp Pro Thr Ala Ile Asn				
35	1170	1175	1180		
	Thr Val Tyr Ser His Val Leu Asn Pro Met Val Ala Ser Arg Ser Tyr				
	1185	1190	1195	1200	
	Met Cys Ile Glu Pro Val Glu Leu His Leu Asp Val His Ala Leu Arg				
	1205	1210	1215		
40	Arg Gly Leu Val Met Pro Val Asp Pro Asp Thr Glu Pro Asn Ala Leu				

	1220	1225	1230
	Leu Val Gln Asp Ser Gly Met Val Pro Val Ser Thr Gln Ile Ser Ile		
	1235	1240	1245
	Val Asn Pro Glu Thr Asn Gln Leu Cys Leu Asn Gly Glu Tyr Gly Glu		
5	1250	1255	1260
	Ile Trp Val Gln Ser Glu Ala Asn Ala Tyr Ser Phe Tyr Met Ser Lys		
	1265	1270	1275
	Glu Arg Leu Asp Ala Glu Arg Phe Asn Gly Arg Thr Ile Asp Gly Asp		1280
	1285	1290	1295
10	Pro Asn Val Arg Tyr Val Arg Thr Gly Asp Leu Gly Phe Leu His Ser		
	1300	1305	1310
	Val Thr Arg Pro Ile Gly Pro Asn Gly Ala Pro Val Asp Met Gln Val		
	1315	1320	1325
	Leu Phe Val Leu Gly Ser Ile Gly Asp Thr Phe Glu Val Asn Gly Leu		
15	1330	1335	1340
	Asn His Phe Ser Met Asp Ile Glu Gln Ser Val Glu Arg Cys His Arg		
	1345	1350	1355
	Asn Ile Val Pro Gly Gly Cys Ala Val Phe Gln Ala Gly Gly Leu Val		
	1365	1370	1375
20	Val Val Val Val Glu Ile Phe Arg Arg Asn Phe Leu Ala Ser Met Val		
	1380	1385	1390
	Pro Val Ile Val Asn Ala Ile Leu Asn Glu His Gln Leu Val Ile Asp		
	1395	1400	1405
	Ile Val Ser Phe Val Gln Lys Gly Asp Phe His Arg Ser Arg Leu Gly		
25	1410	1415	1420
	Glu Lys Gln Arg Gly Lys Ile Leu Ala Gly Trp Val Thr Arg Lys Met		
	1425	1430	1435
	Arg Thr Ile Ala Gln Tyr Ser Ile Arg Asp Pro Asn Gly Gln Asp Ser		
	1445	1450	1455
30	Gln Met Ile Thr Glu Glu Pro Gly Pro Arg Ala Ser Met Thr Gly Ser		
	1460	1465	1470
	Met Leu Gly Arg Met Gly Gly Pro Ala Ser Ile Lys Ala Gly Ser Thr		
	1475	1480	1485
	Arg Ala Pro Ser Leu Met Gly Met Thr Ala Thr Met Asn Asn Leu Ser		
35	1490	1495	1500
	Leu Thr Gln Gln Gln Gln Gln Gln Tyr Gln Gln Pro Gly Met Tyr Ala		
	1505	1510	1515
	Gln Gln Gln Gly Met His Pro Gln Gln Gln His Gln Phe Ser Met Ser		1520
	1525	1530	1535
40	Asn Thr Pro Pro Gln Gly Pro Pro Gln Gly Val Glu Leu His Asp Pro		

10

	1540		1545		1550
	Ser Asp Arg Thr Pro Thr Asp Asn Arg His Ser Phe Leu Ala Asp Pro				
	1555		1560		1565
	Arg Met Gln Asn Gln Gly Gln Met Asn Glu Thr Gly Ala Tyr Glu Pro				
5	1570		1575		1580
	Met Asn Tyr Gln Asn Ala Tyr His Pro His Gln Gln Gln Tyr Glu Ser				
	1585		1590		1595
	Glu Asp Gly Gly Ser Arg Leu Ser Gly Pro Val Pro Asp Val Leu Arg				
	1605		1610		1615
10	Pro Gly Pro Ser Ser Gly Ser Ile Glu Gln His Asp Gln Ala Asn Asn				
	1620		1625		1630
	Asp Asn Asn Met Trp Asn Asn Arg Glu Tyr Tyr Gly Asn Ser Pro Ser				
	1635		1640		1645
	Tyr Ala Gly Gly Tyr Thr Gln Asp Gly Asn Ile His Glu Gln Gln Gln				
15	1650		1655		1660
	His Asp Glu Tyr Thr Ser Asn Ala Ser Tyr Gly Gly Asn Gln Gly Ala				
	1665		1670		1675
	Gly Gly Gly Ser Gly Gly Gly Gly Gly Gly Leu Arg Val Ala Asn Arg Asp				
	1685		1690		1695
20	Ser Ser Asp Ser Glu Gly Ala Asp Asp Asp Ala Trp Arg Arg Asp Ala				
	1700		1705		1710
	Leu Ala Gln Ile Asn Phe Ala Gly Gly Ala Ala Ala Ala Ser Ala Gly				
	1715		1720		1725
	Ala Pro Ala Ala Gly Ala Ser Ser Ser Gln Pro Gly His Ala Gln				
25	1730		1735		1740

<210> 4

<211> 23

<212> DNA

30<213> Artificial Sequence

<220>

<223> Primer

35<400> 4

gcggataaca attttcacaca gga

23

<210> 5

<211> 20

40<212> DNA

11

<213> Artificial Sequence

<220>

<223> Primer

5

<400> 5

aggcccagct gcttctcttg 20

<210> 6

10<211> 24

<212> DNA

<213> Artificial Sequence

<220>

15<223> Primer

<400> 6

actcggaccg gaacggaata acaa 24

20<210> 7

<211> 18

<212> DNA

<213> Artificial Sequence

25<220>

<223> Primer

<400> 7

cggaaggagt gcgaacaa 18

30

<210> 8

<211> 20

<212> DNA

<213> Artificial Sequence

35

<220>

<223> Primer

<400> 8

40gctgcttgca tctggtcttg 20

12

<210> 9

<211> 21

<212> DNA

<213> Artificial Sequence

5

<220>

<223> Primer

<400> 9

10agacccagct gttgccatt g

21

<210> 10

<211> 20

<212> DNA

15<213> Artificial Sequence

<220>

<223> Primer

20<400> 10

cggagacgca aagcctgaga

20

<210> 11

<211> 20

25<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

30

<400> 11

tgccagctgc gtccaagaag

20

<210> 12

35<211> 19

<212> DNA

<213> Artificial Sequence

<220>

40<223> Primer

13

<400> 12

gctagcatgg ccctcacac

19

<210> 13

5<211> 21

<212> DNA

<213> Artificial Sequence

<220>

10<223> Primer

<400> 13

tgtgttgacc tccactagct c

21

15<210> 14

<211> 22

<212> DNA

<213> Artificial Sequence

20<220>

<223> Primer

<400> 14

ctacgggatg cagagggaaa gt

22

25

<210> 15

<211> 21

<212> DNA

<213> Artificial Sequence

30

<220>

<223> Primer

<400> 15

35gccatgatta gcacgatacc c

21

<210> 16

<211> 21

<212> DNA

40<213> Artificial Sequence

14

<220>

<223> Primer

<400> 16

5cgccgtgcat acaactacca a

21

<210> 17

<211> 20

<212> DNA

10<213> Artificial Sequence

<220>

<223> Primer

15<400> 17

tggtggcact acaaagaaca

20

<210> 18

<211> 21

20<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

25

<400> 18

cagcgtcttg aatgggtaga a

21

<210> 19

30<211> 20

<212> DNA

<213> Artificial Sequence

<220>

35<223> Primer

<400> 19

ctgggtagat tcgccttcac

20

40<210> 20

15

<211> 21

<212> DNA

<213> Artificial Sequence

5<220>

<223> Primer

<400> 20

gagcgatcag tgcagaacat t

21

10

<210> 21

<211> 21

<212> DNA

<213> Artificial Sequence

15

<220>

<223> Primer

<400> 21

20cgctgacggtt tgaccatctg a

21

<210> 22

<211> 19

<212> DNA

25<213> Artificial Sequence

<220>

<223> Primer

30<400> 22

gcatatgcaa cgagtcaaa

19

<210> 23

<211> 18

35<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

40

16

<400> 23

acggtgcacc tgttgata

18

<210> 24

5<211> 20

<212> DNA

<213> Artificial Sequence

<220>

10<223> Primer

<400> 24

atgcgcacaa tagcccagta

20

15<210> 25

<211> 21

<212> DNA

<213> Artificial Sequence

20<220>

<223> Primer

<400> 25

ttcaagcaac tgtggcgtag g

21

25

<210> 26

<211> 23

<212> DNA

<213> Artificial Sequence

30

<220>

<223> Primer

<400> 26

35gatcctagcg accgcacacc aac

23

<210> 27

<211> 18

<212> DNA

40<213> Artificial Sequence

17

<220>

<223> Primer

<400> 27

5cctgctgctg gtgcttct

18

<210> 28

<211> 20

<212> DNA

10<213> Artificial Sequence

<220>

<223> Primer

15<400> 28

gagttgcaaa tcgtgacagc

20

<210> 29

<211> 24

20<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

25

<400> 29

tatcagctgt tgttcaatgt tcta

24

<210> 30

30<211> 19

<212> DNA

<213> Artificial Sequence

<220>

35<223> Primer

<400> 30

tggtatccca ttgccattg

19

40<210> 31

18

<211> 20

<212> DNA

<213> Artificial Sequence

5<220>

<223> Primer

<400> 31

aaggacggag attggtggag

20

10

<210> 32

<211> 17

<212> DNA

<213> Artificial Sequence

15

<220>

<223> Primer

<400> 32

20ggagatggcg gtagcga

17

<210> 33

<211> 18

<212> DNA

25<213> Artificial Sequence

<220>

<223> Primer

30<400> 33

gcatggcttg tggaggac

18

<210> 34

<211> 24

35<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

40

<400> 34
agattgtggc tagtatggag gtaa 24

<210> 35
5<211> 17
<212> DNA
<213> Artificial Sequence

<220>
10<223> Primer

<400> 35
gttttcccag tcacgac 17

15<210> 36
<211> 24
<212> DNA
<213> Artificial Sequence

20<220>
<223> Primer

<400> 36
tactactagc ataccagcat acct 24
25

<210> 37
<211> 21
<212> DNA
<213> Artificial Sequence

30
<220>
<223> Primer

<400> 37
35tcaacctcgg aataccaagt c 21

<210> 38
<211> 9
<212> DNA
40<213> Artificial Sequence

20

<220>

<223> Sequence around ATG of ORF1

<400> 38

5caccatgct

9

<210> 39

<211> 9

<212> DNA

10<213> Artificial Sequence

<220>

<223> Fungal consensus

15<400> 39

caccatggc

9

<210> 40

<211> 6003

20<212> DNA

<213> *Fusarium graminearum*

<400> 40

ctcgagggtta gtaaaagatc cccgttttgtt ccacaaatct ccatctccct ctcaatgcct	60
25ttctttggcgc ctcaaccgcg tattttgaag acagtttggt gttgtcgcat gcgacaaaaa	120
atcatcctct caagttttca tcgctgacct gtttcttggt gtaggaagga gatcacacac	180
agaaagggta agctgctttg cgtccagagt acttacaatt gcttctcaat tacttacgcg	240
ccggcagcta caaaagcga cgaactcaac ttttctccca attcctcggg gcacctccac	300
ctcagattgc tgctctcgcc gagcctcagt ctggcctacg catacactcg cccgatgact	360
30ccgaccaccc ttcaggcgat ggccatcgcg ctaccgccta tgccgctctc ggtagcagca	420
gcggtccaat ccagattca ccagactcac ctatgtaccg accgcactct ggttatgctc	480
cttcagaatc accaagacct tctccagcac aacctccacc ttccctgctg cgcccggggg	540
gttctctcgc tggaggatcg accactgctc accgcgactc cctcttcttc tccccctccc	600
atctcgaacc tgaaaccgga acaggtacta tgatgtcggg cgactatgca ttcagaccg	660
35agcagcaagg cacatatggc gaatcccagc atcaacagca ccagttccag caacagcaac	720
agccacagca gcaacagcag tacgatgggc agcagtatga tggacgaact acaacgcttc	780
tcgattcgca aggatacttt tcggattttg cgggacagca gcactatgat cagactcaaa	840
ccgttgagta tgtgggacct cagcagcggg attcttccag cgatgcattc tctccaaccg	900
ccgcaatggc acctccaatg cttacaacca acgacctccc accgccggaa gcgcttgagt	960
40accagctgcc ccttgaccct cgcgaggtac cattcgctat tcaagatccc catgatgatt	1020

	ctacgccaat	gtcaaagttc	gataacatcg	cagctgtact	cagacataga	ggccgaacga	1080
	ttgctaagaa	gccggcatac	tgggtgttgg	atagtaaggg	caaggagatt	gcatcgatta	1140
	cgtgggataa	gctggcatct	agagccgaaa	aggttgcgca	agtcattccgc	gacaaaagct	1200
	ctctgtaccg	gggtgatcgg	gttgctctca	tctaccgcga	ttcagagggtt	attgatttcg	1260
5	ccattgcctt	gctgggatgc	ttcattgctg	gagttgttgc	cggtcccatc	aatgatctgc	1320
	aggactacca	acgcttgaac	cacattctta	ctacaacgca	ggcccatcta	gcgctgacca	1380
	ccgataacaa	cctcaaagcc	tttcaacgag	acattactac	acaaaagttg	acatggccaa	1440
	aggggtgtcg	atgggtggaag	acaaacgagt	ttggcagtta	tcaccccaag	agaaggagg	1500
	atgtcccggc	tttggttgtt	ccgatctgg	catatatcga	gttttcgcgg	gccccactg	1560
10	gagacttgag	aggtgttgtt	ctgagccacc	gaaccattat	gcaccaaattg	gcttgtctta	1620
	gtgcgattat	ttctactatc	ccgggtaatg	gacctggcga	cactttcaac	ccgtctcttc	1680
	gcgacaagaa	tggtcgactt	attggtggcg	gcgcaagcag	cgaaattttg	gtgtcgtacc	1740
	tcgatccccg	tcagggcatt	ggcatgattc	tgagcgtgct	actgaccgtc	tacggcggcc	1800
	acaccactgt	ttggttcgac	aacaaagctg	ttgatgttcc	tggaactgtac	gcccacctcc	1860
15	ttaccaagta	caaatcgacc	atcatgattg	ccgactaccc	aggattgaag	cgagccgcct	1920
	acaactacca	gcaagagcca	atgggtgacc	gaaattttta	gaagggaatg	gagccaaact	1980
	ttcaaattgat	caagctttgc	ttgattgaca	ccttgactgt	agacagcggg	ttccacgaag	2040
	ttttggctga	ccgatggcta	cgaccgttga	gaaaccctcg	tgcccgtgag	gttgtcgcac	2100
	ctatgctttg	tctacctgaa	cacggaggca	tggtgattag	tgtgcgtgac	tggtctaggag	2160
20	gagaagagcg	catgggatgc	ccattaaagc	ttgaacttgg	ggaggatata	gagtctgacg	2220
	aagagaaaga	ggaaacagag	aagccagcag	tttccaatgg	cttttggtagt	ctcttgtcag	2280
	gtggtggcac	agcaacaacc	gaagagaggg	caaagaatga	gcttggcgaa	gtccttttgg	2340
	atcgtgaggc	tctaaagacc	aacgaagtgt	tggttggtggc	cataggtaac	gatgcccgta	2400
	aaagggtgac	ggatgaccca	ggcttggtac	gggtcggttc	ttttggatac	cccatacccg	2460
25	atgccacact	ctccgtcgtc	gatccagaaa	cgggtttact	ggcgtcacca	cattccgtgg	2520
	gtgaaatctg	ggtcgactcc	ccttctcttt	caggtggttt	ctgggcgcag	ccaaagaata	2580
	ctgagctgat	tttccatgct	cgctccttaca	agtttgaccc	aggatgacct	acaccgcagc	2640
	ccgtcgagcc	cgaattcctg	cgaacaggct	tgctgggcac	cgtcatcgag	ggtaaaatct	2700
	ttgttctggg	cctttacgaa	gaccgaattc	gacaaaaggt	tgagtgggtt	gagcatggac	2760
30	acgaactagc	agagtaccgc	tacttctttg	ttcagcacat	cgttgtgagc	attgtcaaga	2820
	acgttccaaa	gatatacgat	tgttcagcct	ttgacgtctt	tgtcaatgac	gaacacctgc	2880
	cagtcgtggg	gctggagtca	gcagctgcgt	caacggcacc	attgacatct	ggaggacctc	2940
	ctcgacaacc	ggatacagct	ctgctagagt	cattggctga	gcgctgcatg	gaggttctca	3000
	tgtcagagca	tcatctgaga	ctgtactgcg	ttatgatcac	agcaccgcac	actttgcctc	3060
35	gagttgttaa	gaacggacga	cgcgaaattg	gtaacatgct	ttgccgtcgg	gagtttgatc	3120
	tcggcaacct	tccatgtgtg	cacgtcaagt	ttggcgtgga	gcatgcagta	cttaacctcc	3180
	ctattggtgt	agaccctata	ggtggatatc	ggtcaccgtt	ggcgtccgat	tctcgtgccg	3240
	aattcttatt	gccagctgac	aagcaatact	ctggtgtcga	caggcgcgaa	gtcgttatcg	3300
	atgaccgtac	ttcaacgccc	ctaaacaatt	tctcttgcac	ttcggatctt	atccaatggc	3360
40	gcgtggcccc	tcaaccagaa	gagctagcgt	actgcacaat	cgatggcaaa	agccgagaag	3420

	gtaagggtgt	aacatggaag	aaattcgaca	ccaaggctgc	ttccgttgcc	atgtacctga	3480
	agaacaaggt	caaggtgagg	ccgggagacc	acatcatcct	catgtacaca	cattcagagg	3540
	agtttgtctt	tgccatccat	gcctgcattt	ccttggggcg	aattgtcatt	cccacgcac	3600
	ccctcgacca	gaaccgattg	aacgaagatg	tcccagcttt	cctgcatatt	gtatctgatt	3660
5	sacaacgtcaa	ggctgtgctg	gtcaacgctg	aggctgatca	tctaatacaag	gtaaagcctg	3720
	tggctagcca	tatcaaacag	tcagcccagg	ttctcaagat	cacgagccct	gccatctaca	3780
	acacaactaa	gccgccaaag	caaagtagtg	gattgaggga	tttgagattc	accattgacc	3840
	ctgcctggat	tcggcctggc	taccccgta	ttgtttggac	ttattggacc	cccgatcaac	3900
	gacgaatttc	agttcagctt	ggacatgaca	ccattatggg	catgtgcaag	gttcaaaagg	3960
10	aaacttgcca	aatgacaagt	tcaagacctg	tgcttggatg	tgtacgaagc	acgactggcc	4020
	taggccttat	tcatacggct	ctgatgggaa	tttatatcgg	aacaccaacc	tacctcctat	4080
	cacctgtcga	gtttgcagcc	aaccccatgt	ctctattcgt	caccttgtcg	agatacaaga	4140
	ttaaggatac	ttatgcgaca	ccacagatgc	ttgatcatgc	catgaactcc	atgcaggcca	4200
	agggctttac	acttcatgaa	cttaagaaca	tgatgatcac	tgccgagagc	cgaccaagag	4260
15	ttgatgtttt	ccaaaaggtc	agacttcact	ttgctggggc	tggtctcgat	agaactgcta	4320
	ttaacacggg	ctattcgcat	gtcctcaacc	ccatggtagc	gtcgcgatct	tatatgtgca	4380
	tcgagcctat	tgagctttgg	ttggacacgc	aagcgcttcg	acgtgggtctg	gttattcctg	4440
	tggaccctga	atcagatcct	ctggccctac	tggtacagga	cagcggtatg	gttccagttt	4500
	caacccaaat	agccatcatc	aaccctgaaa	gcagaataca	ctgcctcgat	ggtgagtatg	4560
20	gtgaaatttg	ggtcgactct	gaagcctgcg	tcaagtcatt	ctatggctcc	aaagacgctt	4620
	ttgacgctga	gcgctttgat	ggccgagctc	ttgacggcga	tcccaacatt	cagtatatcc	4680
	gtaccggaga	cttgggtttc	cttcataatg	ttagtgcacc	tattggccct	aatggtgccc	4740
	aggtggacat	gcaagtgttg	tttgttctcg	gcaacattgg	cgagactttt	gagatcaacg	4800
	gattgagcca	tttcccaatg	gatattgaga	actcggtgga	aaaatgccac	agaaacattg	4860
25	tggcgaatgg	ctggtaagta	taaaatctct	atgtgaagcg	aatatgctaa	caaagtcagt	4920
	gcggtgttcc	aagctggtgg	cttgggtggt	gttctggttg	aagtcaaccg	caagccatac	4980
	ctggcatcga	ttgttcccgt	cattgtcaac	gctatcctca	atgaacacca	aatcattgta	5040
	gatatcgctg	cattcgtcaa	caagggagac	ttcccacggg	ctcgtctagg	agagaagcag	5100
	cgtggcaaga	ttcttggtgg	ctgggttagt	agaaagctga	ggactccttg	ccagttctcg	5160
30	attcgcgata	tggacgccga	atccacagct	ggtgatatga	tggatccttc	tagagcatca	5220
	atggtcagcg	tacgaagcgg	aggcgggtgt	gctcccggat	cttctagttt	gaggaatgtc	5280
	gaacctgcgc	ctcaaactct	ggaggaggaa	catgaccaga	tgactcctcg	tcacgaatac	5340
	gaagcagccc	ctaccatgat	ttctgaactt	cccagcggcc	aagagacacc	gacagggttt	5400
	cagcactcgc	aatacgaaca	cccaccacaa	tcagccgggt	ctcaagcacc	agcccagctg	5460
35	aacctttctc	accagcccga	tcaaggattc	gatatggact	tttcacgata	tagttcagca	5520
	gagcccgatc	acggccctgt	ccacagacgt	ccagtcccag	gccaaagcca	acaacccgag	5580
	cctatgcaag	ggtacgggtc	agcgccgcc	cagatccggc	taccaggtgt	tgatggacga	5640
	gaggagggag	ggttctggtc	acagcaggaa	aagaacgaga	agagtgaaga	agactggaca	5700
	actgatgcca	tgatgcatat	gaatctggca	ggtgatatga	aaccgccacg	atgataatac	5760
40	acaacataag	agcgaagtga	cgaagcggag	tcggagttgg	gaagcattta	gaaacgaata	5820

23

acaaacaatt ggacttgtcg gtctgatggc ctatttactt cattcataga tgaggattgg 5880
 atagtgaata tgtgattgga taaagcctgg gtttgtgagt ttgtgaatgc agtgggtgct 5940
 tgctataagc tgttttattg aggtctttgg aggagtgtct aacaaagatg caaagttact 6000
 agt 6003

5

<210> 41

<211> 1692

10<212> PRT

<213> *Fusarium graminearum*

<400> 41

Met Met Ser Gly Asp Tyr Ala Phe Arg Pro Glu Gln Gln Gly Thr Tyr
 15 1 5 10 15
 Gly Glu Ser Gln His Gln Gln His Gln Phe Gln Gln Gln Gln Gln Pro
 20 25 30
 Gln Gln Gln Gln Gln Tyr Asp Gly Gln Gln Tyr Asp Gly Arg Thr Thr
 35 40 45
 20Thr Leu Leu Asp Ser Gln Gly Tyr Phe Ser Asp Phe Ala Gly Gln Gln
 50 55 60
 His Tyr Asp Gln Thr Gln Thr Val Glu Tyr Val Gly Pro Gln Gln Arg
 65 70 75 80
 Tyr Ser Ser Ser Asp Ala Phe Ser Pro Thr Ala Ala Met Ala Pro Pro
 25 85 90 95
 Met Leu Thr Thr Asn Asp Leu Pro Pro Pro Glu Ala Leu Glu Tyr Gln
 100 105 110
 Leu Pro Leu Asp Pro Arg Glu Val Pro Phe Ala Ile Gln Asp Pro His
 115 120 125
 30Asp Asp Ser Thr Pro Met Ser Lys Phe Asp Asn Ile Ala Ala Val Leu
 130 135 140
 Arg His Arg Gly Arg Thr Ile Ala Lys Lys Pro Ala Tyr Trp Val Leu
 145 150 155 160
 Asp Ser Lys Gly Lys Glu Ile Ala Ser Ile Thr Trp Asp Lys Leu Ala
 35 165 170 175
 Ser Arg Ala Glu Lys Val Ala Gln Val Ile Arg Asp Lys Ser Ser Leu
 180 185 190
 Tyr Arg Gly Asp Arg Val Ala Leu Ile Tyr Arg Asp Ser Glu Val Ile
 195 200 205
 40Asp Phe Ala Ile Ala Leu Leu Gly Cys Phe Ile Ala Gly Val Val Ala

24

	210		215		220	
	Val Pro Ile Asn Asp Leu Gln Asp Tyr Gln Arg Leu Asn His Ile Leu					
	225		230		235	240
	Thr Thr Thr Gln Ala His Leu Ala Leu Thr Thr Asp Asn Asn Leu Lys					
5		245		250		255
	Ala Phe Gln Arg Asp Ile Thr Thr Gln Lys Leu Thr Trp Pro Lys Gly					
		260		265		270
	Val Glu Trp Trp Lys Thr Asn Glu Phe Gly Ser Tyr His Pro Lys Lys					
	275		280		285	
10	Lys Glu Asp Val Pro Ala Leu Val Val Pro Asp Leu Ala Tyr Ile Glu					
	290		295		300	
	Phe Ser Arg Ala Pro Thr Gly Asp Leu Arg Gly Val Val Leu Ser His					
	305		310		315	320
	Arg Thr Ile Met His Gln Met Ala Cys Leu Ser Ala Ile Ile Ser Thr					
15		325		330		335
	Ile Pro Gly Asn Gly Pro Gly Asp Thr Phe Asn Pro Ser Leu Arg Asp					
		340		345		350
	Lys Asn Gly Arg Leu Ile Gly Gly Gly Ala Ser Ser Glu Ile Leu Val					
	355		360		365	
20	Ser Tyr Leu Asp Pro Arg Gln Gly Ile Gly Met Ile Leu Ser Val Leu					
	370		375		380	
	Leu Thr Val Tyr Gly Gly His Thr Thr Val Trp Phe Asp Asn Lys Ala					
	385		390		395	400
	Val Asp Val Pro Gly Leu Tyr Ala His Leu Leu Thr Lys Tyr Lys Ser					
25		405		410		415
	Thr Ile Met Ile Ala Asp Tyr Pro Gly Leu Lys Arg Ala Ala Tyr Asn					
		420		425		430
	Tyr Gln Gln Glu Pro Met Val Thr Arg Asn Phe Lys Lys Gly Met Glu					
	435		440		445	
30	Pro Asn Phe Gln Met Ile Lys Leu Cys Leu Ile Asp Thr Leu Thr Val					
	450		455		460	
	Asp Ser Gly Ser His Glu Val Leu Ala Asp Arg Trp Leu Arg Pro Leu					
	465		470		475	480
	Arg Asn Pro Arg Ala Arg Glu Val Val Ala Pro Met Leu Cys Leu Pro					
35		485		490		495
	Glu His Gly Gly Met Val Ile Ser Val Arg Asp Trp Leu Gly Gly Glu					
		500		505		510
	Glu Arg Met Gly Cys Pro Leu Lys Leu Glu Leu Gly Glu Asp Thr Glu					
	515		520		525	
40	Ser Asp Glu Glu Lys Glu Glu Thr Glu Lys Pro Ala Val Ser Asn Gly					

25

	530		535		540														
	Phe	Gly	Ser	Leu	Leu	Ser	Gly	Gly	Gly	Thr	Ala	Thr	Thr	Glu	Glu	Arg			
	545					550					555					560			
	Ala	Lys	Asn	Glu	Leu	Gly	Glu	Val	Leu	Leu	Asp	Arg	Glu	Ala	Leu	Lys			
5					565				570						575				
	Thr	Asn	Glu	Val	Val	Val	Val	Ala	Ile	Gly	Asn	Asp	Ala	Arg	Lys	Arg			
				580					585					590					
	Val	Thr	Asp	Asp	Pro	Gly	Leu	Val	Arg	Val	Gly	Ser	Phe	Gly	Tyr	Pro			
	595					600							605						
10	Ile	Pro	Asp	Ala	Thr	Leu	Ser	Val	Val	Asp	Pro	Glu	Thr	Gly	Leu	Leu			
	610					615							620						
	Ala	Ser	Pro	His	Ser	Val	Gly	Glu	Ile	Trp	Val	Asp	Ser	Pro	Ser	Leu			
	625					630							635			640			
	Ser	Gly	Gly	Phe	Trp	Ala	Gln	Pro	Lys	Asn	Thr	Glu	Leu	Ile	Phe	His			
15				645					650					655					
	Ala	Arg	Pro	Tyr	Lys	Phe	Asp	Pro	Gly	Asp	Pro	Thr	Pro	Gln	Pro	Val			
				660					665					670					
	Glu	Pro	Glu	Phe	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Thr	Val	Ile	Glu	Gly			
				675					680					685					
20	Lys	Ile	Phe	Val	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Ile	Arg	Gln	Lys	Val			
	690					695								700					
	Glu	Trp	Val	Glu	His	Gly	His	Glu	Leu	Ala	Glu	Tyr	Arg	Tyr	Phe	Phe			
	705					710							715			720			
	Val	Gln	His	Ile	Val	Val	Ser	Ile	Val	Lys	Asn	Val	Pro	Lys	Ile	Tyr			
25				725					730					735					
	Asp	Cys	Ser	Ala	Phe	Asp	Val	Phe	Val	Asn	Asp	Glu	His	Leu	Pro	Val			
				740					745					750					
	Val	Val	Leu	Glu	Ser	Ala	Ala	Ala	Ser	Thr	Ala	Pro	Leu	Thr	Ser	Gly			
				755					760					765					
30	Gly	Pro	Pro	Arg	Gln	Pro	Asp	Thr	Ala	Leu	Leu	Glu	Ser	Leu	Ala	Glu			
	770					775								780					
	Arg	Cys	Met	Glu	Val	Leu	Met	Ser	Glu	His	His	Leu	Arg	Leu	Tyr	Cys			
	785					790							795			800			
	Val	Met	Ile	Thr	Ala	Pro	Asp	Thr	Leu	Pro	Arg	Val	Val	Lys	Asn	Gly			
35				805					810					815					
	Arg	Arg	Glu	Ile	Gly	Asn	Met	Leu	Cys	Arg	Arg	Glu	Phe	Asp	Leu	Gly			
				820					825					830					
	Asn	Leu	Pro	Cys	Val	His	Val	Lys	Phe	Gly	Val	Glu	His	Ala	Val	Leu			
				835					840					845					
40	Asn	Leu	Pro	Ile	Gly	Val	Asp	Pro	Ile	Gly	Gly	Ile	Trp	Ser	Pro	Leu			

	850		855		860														
	Ala	Ser	Asp	Ser	Arg	Ala	Glu	Phe	Leu	Leu	Pro	Ala	Asp	Lys	Gln	Tyr			
	865					870					875					880			
	Ser	Gly	Val	Asp	Arg	Arg	Glu	Val	Val	Ile	Asp	Asp	Arg	Thr	Ser	Thr			
5					885					890					895				
	Pro	Leu	Asn	Asn	Phe	Ser	Cys	Ile	Ser	Asp	Leu	Ile	Gln	Trp	Arg	Val			
				900					905				910						
	Ala	Arg	Gln	Pro	Glu	Glu	Leu	Ala	Tyr	Cys	Thr	Ile	Asp	Gly	Lys	Ser			
		915					920					925							
10	Arg	Glu	Gly	Lys	Gly	Val	Thr	Trp	Lys	Lys	Phe	Asp	Thr	Lys	Val	Ala			
	930					935					940								
	Ser	Val	Ala	Met	Tyr	Leu	Lys	Asn	Lys	Val	Lys	Val	Arg	Pro	Gly	Asp			
	945				950				955				960						
	His	Ile	Ile	Leu	Met	Tyr	Thr	His	Ser	Glu	Glu	Phe	Val	Phe	Ala	Ile			
15				965				970					975						
	His	Ala	Cys	Ile	Ser	Leu	Gly	Ala	Ile	Val	Ile	Pro	Ile	Ala	Pro	Leu			
			980				985				990								
	Asp	Gln	Asn	Arg	Leu	Asn	Glu	Asp	Val	Pro	Ala	Phe	Leu	His	Ile	Val			
		995				1000				1005									
20	Ser	Asp	Tyr	Asn	Val	Lys	Ala	Val	Leu	Val	Asn	Ala	Glu	Val	Asp	His			
	1010					1015					1020								
	Leu	Ile	Lys	Val	Lys	Pro	Val	Ala	Ser	His	Ile	Lys	Gln	Ser	Ala	Gln			
	1025				1030				1035				1040						
	Val	Leu	Lys	Ile	Thr	Ser	Pro	Ala	Ile	Tyr	Asn	Thr	Thr	Lys	Pro	Pro			
25			1045				1050				1055								
	Lys	Gln	Ser	Ser	Gly	Leu	Arg	Asp	Leu	Arg	Phe	Thr	Ile	Asp	Pro	Ala			
			1060				1065				1070								
	Trp	Ile	Arg	Pro	Gly	Tyr	Pro	Val	Ile	Val	Trp	Thr	Tyr	Trp	Thr	Pro			
		1075				1080					1085								
30	Asp	Gln	Arg	Arg	Ile	Ser	Val	Gln	Leu	Gly	His	Asp	Thr	Ile	Met	Gly			
	1090					1095					1100								
	Met	Cys	Lys	Val	Gln	Lys	Glu	Thr	Cys	Gln	Met	Thr	Ser	Ser	Arg	Pro			
	1105				1110				1115				1120						
	Val	Leu	Gly	Cys	Val	Arg	Ser	Thr	Thr	Gly	Leu	Gly	Phe	Ile	His	Thr			
35			1125				1130				1135								
	Ala	Leu	Met	Gly	Ile	Tyr	Ile	Gly	Thr	Pro	Thr	Tyr	Leu	Leu	Ser	Pro			
			1140				1145				1150								
	Val	Glu	Phe	Ala	Ala	Asn	Pro	Met	Ser	Leu	Phe	Val	Thr	Leu	Ser	Arg			
		1155				1160					1165								
40	Tyr	Lys	Ile	Lys	Asp	Thr	Tyr	Ala	Thr	Pro	Gln	Met	Leu	Asp	His	Ala			

	1170	1175	1180	
	Met Asn Ser Met Gln Ala Lys Gly Phe Thr Leu His Glu Leu Lys Asn			
	1185	1190	1195	1200
	Met Met Ile Thr Ala Glu Ser Arg Pro Arg Val Asp Val Phe Gln Lys			
5	1205	1210	1215	
	Val Arg Leu His Phe Ala Gly Ala Gly Leu Asp Arg Thr Ala Ile Asn			
	1220	1225	1230	
	Thr Val Tyr Ser His Val Leu Asn Pro Met Val Ala Ser Arg Ser Tyr			
	1235	1240	1245	
10	Met Cys Ile Glu Pro Ile Glu Leu Trp Leu Asp Thr Gln Ala Leu Arg			
	1250	1255	1260	
	Arg Gly Leu Val Ile Pro Val Asp Pro Glu Ser Asp Pro Leu Ala Leu			
	1265	1270	1275	1280
	Leu Val Gln Asp Ser Gly Met Val Pro Val Ser Thr Gln Ile Ala Ile			
15	1285	1290	1295	
	Ile Asn Pro Glu Ser Arg Ile His Cys Leu Asp Gly Glu Tyr Gly Glu			
	1300	1305	1310	
	Ile Trp Val Asp Ser Glu Ala Cys Val Lys Ser Phe Tyr Gly Ser Lys			
	1315	1320	1325	
20	Asp Ala Phe Asp Ala Glu Arg Phe Asp Gly Arg Ala Leu Asp Gly Asp			
	1330	1335	1340	
	Pro Asn Ile Gln Tyr Ile Arg Thr Gly Asp Leu Gly Phe Leu His Asn			
	1345	1350	1355	1360
	Val Ser Arg Pro Ile Gly Pro Asn Gly Ala Gln Val Asp Met Gln Val			
25	1365	1370	1375	
	Leu Phe Val Leu Gly Asn Ile Gly Glu Thr Phe Glu Ile Asn Gly Leu			
	1380	1385	1390	
	Ser His Phe Pro Met Asp Ile Glu Asn Ser Val Glu Lys Cys His Arg			
	1395	1400	1405	
30	Asn Ile Val Ala Asn Gly Cys Ala Val Phe Gln Ala Gly Gly Leu Val			
	1410	1415	1420	
	Val Val Leu Val Glu Val Asn Arg Lys Pro Tyr Leu Ala Ser Ile Val			
	1425	1430	1435	1440
	Pro Val Ile Val Asn Ala Ile Leu Asn Glu His Gln Ile Ile Val Asp			
35	1445	1450	1455	
	Ile Val Ala Phe Val Asn Lys Gly Asp Phe Pro Arg Ser Arg Leu Gly			
	1460	1465	1470	
	Glu Lys Gln Arg Gly Lys Ile Leu Gly Gly Trp Val Ser Arg Lys Leu			
	1475	1480	1485	
40	Arg Thr Leu Ala Gln Phe Ser Ile Arg Asp Met Asp Ala Glu Ser Thr			

28

1490	1495	1500
Ala Gly Asp Met Met Asp Pro Ser Arg Ala Ser Met Val Ser Val Arg		
1505	1510	1515
Ser Gly Gly Gly Ala Ala Pro Gly Ser Ser Ser Leu Arg Asn Val Glu		
5	1525	1530
Pro Ala Pro Gln Ile Leu Glu Glu Glu His Asp Gln Met Thr Pro Arg		
	1540	1545
His Glu Tyr Glu Ala Ala Pro Thr Met Ile Ser Glu Leu Pro Asp Gly		
	1555	1560
10Gln Glu Thr Pro Thr Gly Phe Gln His Ser Gln Tyr Glu His Pro Pro		
	1570	1575
Gln Ser Ala Gly Ser Gln Ala Pro Ala Gln Leu Asn Leu Ser His Gln		
1585	1590	1595
Pro Asp Gln Gly Phe Asp Met Asp Phe Ser Arg Tyr Ser Ser Ala Glu		
15	1605	1610
Pro Asp His Gly Pro Val His Arg Arg Pro Val Pro Gly Gln Ala Gln		
	1620	1625
Gln Pro Glu Pro Met Gln Gly Tyr Gly Gln Ala Pro Pro Gln Ile Arg		
	1635	1640
20Leu Pro Gly Val Asp Gly Arg Glu Glu Gly Gly Phe Trp Ser Gln Gln		
	1650	1655
Glu Lys Asn Glu Lys Ser Glu Glu Asp Trp Thr Thr Asp Ala Met Met		
1665	1670	1675
His Met Asn Leu Ala Gly Asp Met Lys Pro Pro Arg		
25	1685	1690

<210> 42

<211> 2369

<212> DNA

30<213> Alternaria solani

<400> 42

aagaagaaag ggccgaccga gttgaccgaa atattgctag ataaggaagc actgaagctg	60
aacgaagttg ttgttttggc cattggagag gaagtgagca agcgtgtcaa cgaaccggc	120
35actatgagag tcggtgcttt tggctaccg ataccagatg cgacgctggc cgtcgtcgat	180
ccggaaacta atcttttgtg ttcacctat tccataggag agatctgggt agactcgcca	240
tcattgtccg gagggttttg gcagctgcag aagcacactg agactatttt ccacgctcgg	300
ccatatcggt tcgtagagg cagcccaacc ccgcaactac tcgaactgga gtttctacgc	360
actggactgc tcggatgcgt ggtagaaggc aaaatcttcg tattaggcct gtacgaggac	420
40cggattaggc agcgcgttga atgggtagag cacggtcagc tagaagccga acataggtat	480

```

ttcttcgtgc agcatcttgt caccagcatt atgaaagctg ttccaaagat ttacgactgg 540
taagtgctat cgaatctctg ggtaatcaac ctaacattgc gcagctcgtc ttctgattcc 600
tatgtcaacg gcgaatactt accaatcatc cttatcgaga cacaggccgc atcaactgct 660
cccacaaatc caggcggggc accacaacaa cttgacattc ctttcctaga ctctctttct 720
5gagcgtatgta tggaggtact gtatcaagaa caccaccttc ggggtgattg tgtgatgac 780
actgcaccga acacactccc gcgagtcac cagaacgggc gacgagaaat tggaaacatg 840
ctttgccgga gagaatttga caatggctcg ctaccctgcg ttacagtcac gtttgccgtc 900
gagaggctcg tccagaatat tgcgctaggt gatgaccctg ctggcggcat gtggctctac 960
gaggcgtcga tggcacgcca gcagttcctg atgcttcaag ataagcagta ctctggagta 1020
10gatcacagag aagtcgttat tgacgacaga acgtcgacgc cgctcaacca gttctccaac 1080
attcatgacc ttatgcaatg gcgctgacaa cgacaagctg aagagctcgc ctactgcacg 1140
gtagatggtc gaggtaaaga gggcaaaggc gtcaactgga agaagttcga ccagaaggtc 1200
gcagggtgtcg ccatgtacct gaagaacaag gtcaagggtc agactgggtga ccacctgctc 1260
ttgatgtaca cccactcgga agactttgtc tatgccgtac acgctgtgtt cgctccttga 1320
15gctgtgtgta taccatggc accaatcgac cagaacaggc taaatgaaga cgcgcccgca 1380
ctactacata tcattgctga cttcaaggtc aaggctatcc tcgtcaatgc tggcgtagac 1440
cacctgatga aggtcaagca agtatcgag cacatcaaac agtcagcagt cattctcaag 1500
atcaacgtac cgaataccta taacaccaca aaaccaccta agcagtctag tggttgccgc 1560
gatcttaagc tcacaatacg acctgcttgg atacaatctg gtttcctgt tctagtatgg 1620
20acatactgga cacctgacca gagacgcata gctgtgcaat taggtcatag ccaaatcatg 1680
gcgctatgca aagttcagaa agaaacgtgc cagatgacga gcacacggcc cgctccttga 1740
tgtgttcgta gcacgatcg tcttggttc atacacacct gtgttatggg tatcttcctc 1800
gcagcgccaa cttaccttgt gtcacctgtc gattttgcgc aaaacccgaa catcctcttc 1860
cagaccatgt cgagatacaa gatcaaggac gcgtatgcga ccagccaaat gctggaccac 1920
25gctattgcac gaggtgctgg caagaacatg gctctgcacg agctcaagaa cctcatgatc 1980
gcgactgacg gtcggccgcg cgtagacgtc tgtaagtgtt gcgatcctgt ataagcatct 2040
gaaatctaata tcttgataga ccagcgtgtg cgagtacact tctcgccagc aagtttggac 2100
cgaacggcaa tcaatactgt ttactcacac gtactgaatc ctatggctgc atcgcggtca 2160
tacatgtgca tcgaacccat agaactacat ctcgatgtcg gtgcccttcg aagaggctctc 2220
30atcatgcctg tcgaccaga cagggaacct ggtgctctct tagtccagga ctcggttatg 2280
gtaccagtta gtacacaaat ttcaatcgtg aatccagaga caaacagct ttgcctagtc 2340
ggcgagtatg gcgaaatctg ggtccaacc 2369

```

<210> 43

35<211> 758

<212> PRT

<213> *Alternaria solani*

<400> 43

40Lys Lys Lys Gly Pro Thr Glu Leu Thr Glu Ile Leu Leu Asp Lys Glu

30

1	5	10	15
Ala Leu Lys Leu Asn Glu Val Val Val Leu Ala Ile Gly Glu Glu Val			
	20	25	30
Ser Lys Arg Val Asn Glu Pro Gly Thr Met Arg Val Gly Ala Phe Gly			
5	35	40	45
Tyr Pro Ile Pro Asp Ala Thr Leu Ala Val Val Asp Pro Glu Thr Asn			
	50	55	60
Leu Leu Cys Ser Pro Tyr Ser Ile Gly Glu Ile Trp Val Asp Ser Pro			
65	70	75	80
10Ser Leu Ser Gly Gly Phe Trp Gln Leu Gln Lys His Thr Glu Thr Ile			
	85	90	95
Phe His Ala Arg Pro Tyr Arg Phe Val Glu Gly Ser Pro Thr Pro Gln			
	100	105	110
Leu Leu Glu Leu Glu Phe Leu Arg Thr Gly Leu Leu Gly Cys Val Val			
15	115	120	125
Glu Gly Lys Ile Phe Val Leu Gly Leu Tyr Glu Asp Arg Ile Arg Gln			
	130	135	140
Arg Val Glu Trp Val Glu His Gly Gln Leu Glu Ala Glu His Arg Tyr			
145	150	155	160
20Phe Phe Val Gln His Leu Val Thr Ser Ile Met Lys Ala Val Pro Lys			
	165	170	175
Ile Tyr Asp Cys Ser Ser Phe Asp Ser Tyr Val Asn Gly Glu Tyr Leu			
	180	185	190
Pro Ile Ile Leu Ile Glu Thr Gln Ala Ala Ser Thr Ala Pro Thr Asn			
25	195	200	205
Pro Gly Gly Pro Pro Gln Gln Leu Asp Ile Pro Phe Leu Asp Ser Leu			
	210	215	220
Ser Glu Arg Cys Met Glu Val Leu Tyr Gln Glu His His Leu Arg Val			
225	230	235	240
30Tyr Cys Val Met Ile Thr Ala Pro Asn Thr Leu Pro Arg Val Ile Lys			
	245	250	255
Asn Gly Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe Asp			
	260	265	270
Asn Gly Ser Leu Pro Cys Val His Val Lys Phe Gly Val Glu Arg Ser			
35	275	280	285
Val Gln Asn Ile Ala Leu Gly Asp Asp Pro Ala Gly Gly Met Trp Ser			
	290	295	300
Tyr Glu Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys			
305	310	315	320
40Gln Tyr Ser Gly Val Asp His Arg Glu Val Val Ile Asp Asp Arg Thr			

31

				325					330				335			
	Ser	Thr	Pro	Leu	Asn	Gln	Phe	Ser	Asn	Ile	His	Asp	Leu	Met	Gln	Trp
				340					345				350			
	Arg	Val	Gln	Arg	Gln	Ala	Glu	Glu	Leu	Ala	Tyr	Cys	Thr	Val	Asp	Gly
5			355					360					365			
	Arg	Gly	Lys	Glu	Gly	Lys	Gly	Val	Asn	Trp	Lys	Lys	Phe	Asp	Gln	Lys
			370				375					380				
	Val	Ala	Gly	Val	Ala	Met	Tyr	Leu	Lys	Asn	Lys	Val	Lys	Gly	Gln	Thr
	385				390					395					400	
10	Gly	Asp	His	Leu	Leu	Leu	Met	Tyr	Thr	His	Ser	Glu	Asp	Phe	Val	Tyr
				405					410				415			
	Ala	Val	His	Ala	Cys	Phe	Val	Leu	Gly	Ala	Val	Cys	Ile	Pro	Met	Ala
				420					425				430			
	Pro	Ile	Asp	Gln	Asn	Arg	Leu	Asn	Glu	Asp	Ala	Pro	Ala	Leu	Leu	His
15			435					440					445			
	Ile	Ile	Ala	Asp	Phe	Lys	Val	Lys	Ala	Ile	Leu	Val	Asn	Ala	Gly	Val
			450				455					460				
	Asp	His	Leu	Met	Lys	Val	Lys	Gln	Val	Ser	Gln	His	Ile	Lys	Gln	Ser
	465				470					475				480		
20	Ala	Val	Ile	Leu	Lys	Ile	Asn	Val	Pro	Asn	Thr	Tyr	Asn	Thr	Thr	Lys
				485					490				495			
	Pro	Pro	Lys	Gln	Ser	Ser	Gly	Cys	Arg	Asp	Leu	Lys	Leu	Thr	Ile	Arg
			500						505				510			
	Pro	Ala	Trp	Ile	Gln	Ser	Gly	Phe	Pro	Val	Leu	Val	Trp	Thr	Tyr	Trp
25			515					520					525			
	Thr	Pro	Asp	Gln	Arg	Arg	Ile	Ala	Val	Gln	Leu	Gly	His	Ser	Gln	Ile
			530				535					540				
	Met	Ala	Leu	Cys	Lys	Val	Gln	Lys	Glu	Thr	Cys	Gln	Met	Thr	Ser	Thr
	545				550					555				560		
30	Arg	Pro	Val	Leu	Gly	Cys	Val	Arg	Ser	Thr	Ile	Gly	Leu	Gly	Phe	Ile
				565					570				575			
	His	Thr	Cys	Val	Met	Gly	Ile	Phe	Leu	Ala	Ala	Pro	Thr	Tyr	Leu	Val
			580				585					590				
	Ser	Pro	Val	Asp	Phe	Ala	Gln	Asn	Pro	Asn	Ile	Leu	Phe	Gln	Thr	Met
35			595					600				605				
	Ser	Arg	Tyr	Lys	Ile	Lys	Asp	Ala	Tyr	Ala	Thr	Ser	Gln	Met	Leu	Asp
			610				615					620				
	His	Ala	Ile	Ala	Arg	Gly	Ala	Gly	Lys	Asn	Met	Ala	Leu	His	Glu	Leu
	625				630				635				640			
40	Lys	Asn	Leu	Met	Ile	Ala	Thr	Asp	Gly	Arg	Pro	Arg	Val	Asp	Val	Tyr

32

	645	650	655
	Gln Arg Val Arg Val His Phe Ser Pro Ala Ser Leu Asp Arg Thr Ala		
	660	665	670
	Ile Asn Thr Val Tyr Ser His Val Leu Asn Pro Met Val Ala Ser Arg		
5	675	680	685
	Ser Tyr Met Cys Ile Glu Pro Ile Glu Leu His Leu Asp Val Gly Ala		
	690	695	700
	Leu Arg Arg Gly Leu Ile Met Pro Val Asp Pro Asp Thr Glu Pro Gly		
	705	710	715
10	Ala Leu Leu Val Gln Asp Ser Gly Met Val Pro Val Ser Thr Gln Ile		
	725	730	735
	Ser Ile Val Asn Pro Glu Thr Asn Gln Leu Cys Leu Val Gly Glu Tyr		
	740	745	750
	Gly Glu Ile Trp Val Gln		
15	755		

<210> 44

<211> 2320

<212> DNA

20<213> Pyrenophora teres

<400> 44

aaaaagaagg	ggcctacgga	gttgaccgag	atattgctag	ataaggaagc	gctcaagatg	60
aacgatgttg	tggtccttgc	aataggagaa	gaggccagta	aacgtgcgaa	tgagcctggc	120
25acaatgcgag	ttggcgcttt	tggataccca	ataccagatg	cgacgctagc	cgtcgtagat	180
ccagagacga	atctcttgtg	ttcaccctac	tcgataggag	agatttgggt	agactcacct	240
tcattgtctg	gtggtttctg	gcaattgcag	aagcacactg	aaactatatt	tcacgcccgc	300
ccataccgct	ttgtggaggg	cagtcctacc	ccgcagttgc	ttgagcttga	gtttctccgg	360
acaggcttac	tcggattcgt	cgtagagggc	aaggctctta	tccttggtct	ctatgaagat	420
30cgcatcaggc	agcgcgttga	atgggtagaa	catggctcagc	tggaagctga	acacagatac	480
ttcttcgtgc	agcacctcgt	caccagtatc	atgaaggctg	ttcccaagat	ctacgactgg	540
taagtcttct	catgttttag	atgagcgttc	taacactatg	cagctcatct	ttcgactcgt	600
acgtcaatgg	cgaatacctg	cctatcatcc	tcatcgagac	acaggctgca	tcgacagccc	660
ctacgaaccc	tggtggaccg	ccacagcaac	tcgacatccc	cttcctagac	tcactgtctg	720
35agcgatgcat	ggaagtgttg	tatcaagaac	accatctgcg	agtatactgc	gtcatgatca	780
cagcgccaaa	cacattacca	cgagttgtta	agaatggctg	acgagaaatt	ggcaacatgc	840
tctgtcgaag	agaatttgat	aatggctcat	tacctgtgtg	ccacgtcaag	tttgggtgtg	900
agaggctcag	tctcaacatc	gcgttgggtg	atgacccctc	cggaggcatg	tggtcatatg	960
aagcctcgat	ggcgcgtcag	cagttcttga	tgctccaaga	caagcagtat	tctggagtag	1020
40atcaccgcga	agtcgtcatg	gatgacagaa	catcgacacc	tctcaaccaa	ttctccaaca	1080


```

ttcacgacct catgcaatgg cgcgtatcac ggcaggctga agagctcgca tattgcacag 1140
tcgacggctcg aggcaaagaa ggcaagggcg tcaactggaa gaagtctgac cagaaagttg 1200
cgggtgtcgc aatgtacctg aagaacaagg tcaaagtgca aaccggcgat catctgcttc 1260
tgatgtatac gcactcggaa gactttgtat atgcggtaca tgcattgctt gtgcttggcg 1320
5ctgtatgcat accaatggca ccaatcgacc agaaccgatt gaatgaggat gcacctgcat 1380
tgctgcacat ccttgacagc ttcaagggtca aggccatcct cgtcaatgcc gatgtggatc 1440
atctcatgaa ggtcaagcaa gtatcgcagc acatcaaaca atcagcagcc atcttcaaga 1500
tcaacgtgcc gcacacttac aacacaacca agccacctaa gcagtcgagt ggttgtcggg 1560
atctcaagct cacaatacgg cctgcctggg tacagcctgg tttcccagtt cttgtatgga 1620
10catactggac tccagatcaa cgcgtatag ccgtacaact aggtcatagc caaatcatgg 1680
cactaggcaa ggtccagaag gagacttgct aaatgacaag tacaaggcca gtcctaggat 1740
gtgtacggag taccatcgga cttggcttca ttcatacctg catcatgggc atcttccttg 1800
ccgcacccac ttacctcgtg tcgcctgtcg actttgcaca aaatccaaac atactcttcc 1860
agacgttatc aagatacaag atcaagaatg cgtacgcaac cagtcaaag ttggatcacg 1920
15ctattgcccg tggggctgga aagaacatgg ccctgcacga actcaagaat ctcatgattg 1980
cgactgatgg taggccgcgt gttgatgttt accagagagt gcgcgtacac ttttcaccag 2040
caagcttgga cgggacagcg attaacacag tctactctca cgtgctcaac ccaatggtag 2100
catcgcgatc atacatgtgc atcgagccaa tagaactgca tctcgacgtc aacgctcttc 2160
gaagaggctc gatcatgccc gtcgaccag ataccgagcc tggcgctcta atgggtccagg 2220
20actctggtat ggtgccagtc tccacacaaa tagcaattgt gaaccagag acaaaccagc 2280
tttgcttggg tggcgaatat ggcgaaatct gggttcaatc 2320

```

<210> 45

<211> 758

25<212> PRT

<213> Pyrenophora teres

<400> 45

```

Lys Lys Lys Gly Pro Thr Glu Leu Thr Glu Ile Leu Leu Asp Lys Glu
30 1          5          10          15
Ala Leu Lys Met Asn Asp Val Val Val Leu Ala Ile Gly Glu Glu Ala
          20          25          30
Ser Lys Arg Ala Asn Glu Pro Gly Thr Met Arg Val Gly Ala Phe Gly
          35          40          45
35Tyr Pro Ile Pro Asp Ala Thr Leu Ala Val Val Asp Pro Glu Thr Asn
          50          55          60
Leu Leu Cys Ser Pro Tyr Ser Ile Gly Glu Ile Trp Val Asp Ser Pro
65          70          75          80
Ser Leu Ser Gly Gly Phe Trp Gln Leu Gln Lys His Thr Glu Thr Ile
40          85          90          95

```

34

Phe His Ala Arg Pro Tyr Arg Phe Val Glu Gly Ser Pro Thr Pro Gln
 100 105 110
 Leu Leu Glu Leu Glu Phe Leu Arg Thr Gly Leu Leu Gly Phe Val Val
 115 120 125
 5Glu Gly Lys Val Phe Ile Leu Gly Leu Tyr Glu Asp Arg Ile Arg Gln
 130 135 140
 Arg Val Glu Trp Val Glu His Gly Gln Leu Glu Ala Glu His Arg Tyr
 145 150 155 160
 Phe Phe Val Gln His Leu Val Thr Ser Ile Met Lys Ala Val Pro Lys
 10 165 170 175
 Ile Tyr Asp Cys Ser Ser Phe Asp Ser Tyr Val Asn Gly Glu Tyr Leu
 180 185 190
 Pro Ile Ile Leu Ile Glu Thr Gln Ala Ala Ser Thr Ala Pro Thr Asn
 195 200 205
 15Pro Gly Gly Pro Pro Gln Gln Leu Asp Ile Pro Phe Leu Asp Ser Leu
 210 215 220
 Ser Glu Arg Cys Met Glu Val Leu Tyr Gln Glu His His Leu Arg Val
 225 230 235 240
 Tyr Cys Val Met Ile Thr Ala Pro Asn Thr Leu Pro Arg Val Val Lys
 20 245 250 255
 Asn Gly Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe Asp
 260 265 270
 Asn Gly Ser Leu Pro Cys Val His Val Lys Phe Gly Val Glu Arg Ser
 275 280 285
 25Val Leu Asn Ile Ala Leu Gly Asp Asp Pro Ser Gly Gly Met Trp Ser
 290 295 300
 Tyr Glu Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys
 305 310 315 320
 Gln Tyr Ser Gly Val Asp His Arg Glu Val Val Met Asp Asp Arg Thr
 30 325 330 335
 Ser Thr Pro Leu Asn Gln Phe Ser Asn Ile His Asp Leu Met Gln Trp
 340 345 350
 Arg Val Ser Arg Gln Ala Glu Glu Leu Ala Tyr Cys Thr Val Asp Gly
 355 360 365
 35Arg Gly Lys Glu Gly Lys Gly Val Asn Trp Lys Lys Phe Asp Gln Lys
 370 375 380
 Val Ala Gly Val Ala Met Tyr Leu Lys Asn Lys Val Lys Val Gln Thr
 385 390 395 400
 Gly Asp His Leu Leu Leu Met Tyr Thr His Ser Glu Asp Phe Val Tyr
 40 405 410 415

35

Ala Val His Ala Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala
 420 425 430
 Pro Ile Asp Gln Asn Arg Leu Asn Glu Asp Ala Pro Ala Leu Leu His
 435 440 445
 5Ile Leu Ala Asp Phe Lys Val Lys Ala Ile Leu Val Asn Ala Asp Val
 450 455 460
 Asp His Leu Met Lys Val Lys Gln Val Ser Gln His Ile Lys Gln Ser
 465 470 475 480
 Ala Ala Ile Phe Lys Ile Asn Val Pro His Thr Tyr Asn Thr Thr Lys
 10 485 490 495
 Pro Pro Lys Gln Ser Ser Gly Cys Arg Asp Leu Lys Leu Thr Ile Arg
 500 505 510
 Pro Ala Trp Val Gln Pro Gly Phe Pro Val Leu Val Trp Thr Tyr Trp
 515 520 525
 15Thr Pro Asp Gln Arg Arg Ile Ala Val Gln Leu Gly His Ser Gln Ile
 530 535 540
 Met Ala Leu Gly Lys Val Gln Lys Glu Thr Cys Gln Met Thr Ser Thr
 545 550 555 560
 Arg Pro Val Leu Gly Cys Val Arg Ser Thr Ile Gly Leu Gly Phe Ile
 20 565 570 575
 His Thr Cys Ile Met Gly Ile Phe Leu Ala Ala Pro Thr Tyr Leu Val
 580 585 590
 Ser Pro Val Asp Phe Ala Gln Asn Pro Asn Ile Leu Phe Gln Thr Leu
 595 600 605
 25Ser Arg Tyr Lys Ile Lys Asn Ala Tyr Ala Thr Ser Gln Met Leu Asp
 610 615 620
 His Ala Ile Ala Arg Gly Ala Gly Lys Asn Met Ala Leu His Glu Leu
 625 630 635 640
 Lys Asn Leu Met Ile Ala Thr Asp Gly Arg Pro Arg Val Asp Val Tyr
 30 645 650 655
 Gln Arg Val Arg Val His Phe Ser Pro Ala Ser Leu Asp Arg Thr Ala
 660 665 670
 Ile Asn Thr Val Tyr Ser His Val Leu Asn Pro Met Val Ala Ser Arg
 675 680 685
 35Ser Tyr Met Cys Ile Glu Pro Ile Glu Leu His Leu Asp Val Asn Ala
 690 695 700
 Leu Arg Arg Gly Leu Ile Met Pro Val Asp Pro Asp Thr Glu Pro Gly
 705 710 715 720
 Ala Leu Met Val Gln Asp Ser Gly Met Val Pro Val Ser Thr Gln Ile
 40 725 730 735

Ala Ile Val Asn Pro Glu Thr Asn Gln Leu Cys Leu Val Gly Glu Tyr

740

745

750

Gly Glu Ile Trp Val Gln

755

5

<210> 46

<211> 2435

<212> DNA

<213> *Coccidioides immitis*

10

<400> 46

ggggtggaat	ggtggaagac	aaacgagttt	ggtagctatc	accctaagcg	aaaggatgag	60
atgccccccc	tagccgtccc	ggatttggca	tacatcgagt	ttgcgagggc	tcccactggc	120
gattttgcggg	gagtggtgat	gagccaccgc	accatcatgc	atcaaatgtg	ctgcatgtct	180
15gcgatagtat	ctacgattcc	caccgattcc	aataatagcg	ggaaacccgt	gccaagacct	240
cacggcgaaa	tcctgatgag	ttatctcgat	cctagacaag	gcattggcat	gatccttggg	300
gttctcctta	cggctctatg	tggaataact	actgtttggc	tagagtccct	agcggttgaa	360
actcccgggc	tttatgctag	tttgatcacc	aagtacaggg	ctgctctgct	ggcagcagat	420
tacccgggcc	ttaagagggc	cgtgtacaat	taccagcaag	atccgatggc	gacaagaaat	480
20ttaagaaga	attcagagcc	aaacttctca	agcttgaagt	tgtgtcttat	agatacttta	540
actgtcgact	gcgaattcca	tgaaatcctc	gccgacagat	ggttaaggcc	cttgcggaat	600
ccgcgggctc	gcgaactagt	tacgcccattg	ctgtgccttc	cagaacacgg	tggcatgggt	660
atcagttttac	gtgactggct	tgaggcgag	gagcgtatgg	ggtgcccttt	gaaacatgaa	720
gtactgccac	cggaaaagca	gaaagacaag	tccgaagggtg	agaaaaaaga	agaagagaag	780
25ggcggagagc	caaaggcgac	gttcgggagc	agcttgattg	gtggttctgc	ggcgccgata	840
cgaaaagaag	gcccccgga	cgaccttggg	gagggtactac	ttgacaaaaga	agccttgaaa	900
aacaacgaaa	tttgtgatatt	agcaattggg	gaggaggcaa	gaaggctggc	tgacacaaca	960
ccaaatgctg	tcagggttgg	tgcatttggg	tatcccatc	cagatgcaac	gttagcgatc	1020
gttgatccag	agactgggtt	gctgtgcacg	cctaattgtg	ttgggtgagat	atgggttgat	1080
30tcaccttcat	tgtcaggagg	attctggggc	cttcccaaac	aaacggagtc	catcttccat	1140
gcccgtccct	accgatttca	gggagggggg	cccacacctg	taatcgtgga	gcctgaattc	1200
ttgcgaacag	ggcttcttgg	ctgtgttatt	gagggtcaaa	tattcgtgct	tggctctctac	1260
gaagatcgct	tgcgccaaaa	agttgaatgg	gttgagcatg	gcgtagaagt	tgacagcac	1320
cgatatttct	tcgtgcaaca	tctgattctc	agtattatga	agaacgtgcc	caaaattcac	1380
35gactgctctg	cctttgacgt	cttcgtcaac	gaggagcacc	tgccagtcgt	tgtcttggag	1440
tcgtacactg	cctcaacagc	accagtagct	tcagggaat	ccccacgaca	gctggacgtt	1500
cctcttttgg	actccttggc	tgagaaatgc	atgggagtg	tataccaaga	acatcatctt	1560
cgcgtttatt	gtgtcatgat	cactgccccg	aataccttgc	ctagagttct	taaaaatggg	1620
cgccaagaga	ttggcaacat	gctatgtcga	aaagaatttg	ataatgggtc	gctgccatgc	1680
40gagcacgtta	aattcagcgt	tgagcggtcg	gttctgagtc	ttccaattgg	cgtggatccc	1740

37

```

gttggaggaa tttggtctgt tccatcttca gctgctaggg aggatgccct cgccatgcag      1800
gaaaagcaat attcaggagt cgatttgagg gacggttatta tggatgatcg cacctctacg      1860
ccattgaata attttaacag tatcggtgat ttacttcagt ggcgtgtttc tcgccagggc      1920
gaggaacttt gttattgctc tatcgacggg cgtggcagag aaggcaaggg tatcacatgg      1980
5aagaaattcg attctaaagt tgcagctgtg gctgcgtatt tgaaaaataa agtgaaactc      2040
cgccccggcg accatgttat tctcatgtat acgcactcgg aagagtacgt attcgccgta      2100
catgcttgct tctgcctggg cttggtagcc attcccattt cccagttga ccagaaccga      2160
ctatccgaag atgcgccggc ttactccat gtcattgtcg atttccgtgt aaaagccata      2220
cttgtaacg gcgaagtcaa tgacttactg aaacagaaaa tcgtatctca gcatatcaag      2280
10cagtctgctc atgttggtccg cagcagcgtt ccaagtgtat acaatacgtc gaagccccc      2340
aagcaatcgc acggttgccg ccatctagga ttactatga atccccaatg gttgaattct      2400
aagcagccag cagtgatttg gacctactgg acgcc                                     2435

```

<210> 47

15<211> 812

<212> PRT

<213> *Coccidioides immitis*

<400> 47

```

20Gly Val Glu Trp Trp Lys Thr Asn Glu Phe Gly Ser Tyr His Pro Lys
   1             5             10             15
Arg Lys Asp Glu Met Pro Pro Leu Ala Val Pro Asp Leu Ala Tyr Ile
           20           25           30
Glu Phe Ala Arg Ala Pro Thr Gly Asp Leu Arg Gly Val Val Met Ser
25           35           40           45
His Arg Thr Ile Met His Gln Met Cys Cys Met Ser Ala Ile Val Ser
           50           55           60
Thr Ile Pro Thr Asp Ser Asn Asn Ser Gly Lys Pro Val Pro Arg Pro
65           70           75           80
30His Gly Glu Ile Leu Met Ser Tyr Leu Asp Pro Arg Gln Gly Ile Gly
           85           90           95
Met Ile Leu Gly Val Leu Leu Thr Val Tyr Ala Gly Asn Thr Thr Val
           100          105          110
Trp Leu Glu Ser Leu Ala Val Glu Thr Pro Gly Leu Tyr Ala Ser Leu
35           115          120          125
Ile Thr Lys Tyr Arg Ala Ala Leu Leu Ala Ala Asp Tyr Pro Gly Leu
           130          135          140
Lys Arg Ala Val Tyr Asn Tyr Gln Gln Asp Pro Met Ala Thr Arg Asn
145           150           155           160
40Phe Lys Lys Asn Ser Glu Pro Asn Phe Ser Ser Leu Lys Leu Cys Leu

```

38

				165					170					175			
	Ile	Asp	Thr	Leu	Thr	Val	Asp	Cys	Glu	Phe	His	Glu	Ile	Leu	Ala	Asp	
				180					185					190			
	Arg	Trp	Leu	Arg	Pro	Leu	Arg	Asn	Pro	Arg	Ala	Arg	Glu	Leu	Val	Thr	
5			195					200					205				
	Pro	Met	Leu	Cys	Leu	Pro	Glu	His	Gly	Gly	Met	Val	Ile	Ser	Leu	Arg	
			210					215					220				
	Asp	Trp	Leu	Gly	Gly	Glu	Glu	Arg	Met	Gly	Cys	Pro	Leu	Lys	His	Glu	
	225					230				235					240		
10	Val	Leu	Pro	Pro	Glu	Lys	Gln	Lys	Asp	Lys	Ser	Glu	Gly	Glu	Lys	Lys	
				245						250					255		
	Glu	Glu	Glu	Lys	Gly	Gly	Glu	Pro	Lys	Ala	Thr	Phe	Gly	Ser	Ser	Leu	
				260					265					270			
	Ile	Gly	Gly	Ser	Ala	Ala	Pro	Ile	Arg	Lys	Glu	Gly	Pro	Arg	Asn	Asp	
15			275					280					285				
	Leu	Gly	Glu	Val	Leu	Leu	Asp	Lys	Glu	Ala	Leu	Lys	Asn	Asn	Glu	Ile	
			290					295					300				
	Val	Ile	Leu	Ala	Ile	Gly	Glu	Glu	Ala	Arg	Arg	Leu	Ala	Asp	Thr	Thr	
	305					310				315					320		
20	Pro	Asn	Ala	Val	Arg	Val	Gly	Ala	Phe	Gly	Tyr	Pro	Ile	Pro	Asp	Ala	
				325						330					335		
	Thr	Leu	Ala	Ile	Val	Asp	Pro	Glu	Thr	Gly	Leu	Leu	Cys	Thr	Pro	Asn	
				340					345					350			
	Val	Val	Gly	Glu	Ile	Trp	Val	Asp	Ser	Pro	Ser	Leu	Ser	Gly	Gly	Phe	
25			355					360					365				
	Trp	Ala	Leu	Pro	Lys	Gln	Thr	Glu	Ser	Ile	Phe	His	Ala	Arg	Pro	Tyr	
			370					375				380					
	Arg	Phe	Gln	Gly	Gly	Gly	Pro	Thr	Pro	Val	Ile	Val	Glu	Pro	Glu	Phe	
	385					390				395					400		
30	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Cys	Val	Ile	Glu	Gly	Gln	Ile	Phe	Val	
				405						410					415		
	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Leu	Arg	Gln	Lys	Val	Glu	Trp	Val	Glu	
				420					425					430			
	His	Gly	Val	Glu	Val	Ala	Glu	His	Arg	Tyr	Phe	Phe	Val	Gln	His	Leu	
35			435					440					445				
	Ile	Leu	Ser	Ile	Met	Lys	Asn	Val	Pro	Lys	Ile	His	Asp	Cys	Ser	Ala	
			450					455					460				
	Phe	Asp	Val	Phe	Val	Asn	Glu	Glu	His	Leu	Pro	Val	Val	Val	Leu	Glu	
	465					470				475					480		
40	Ser	Tyr	Thr	Ala	Ser	Thr	Ala	Pro	Val	Ala	Ser	Gly	Gln	Ser	Pro	Arg	

39

					485					490					495	
	Gln	Leu	Asp	Val	Pro	Leu	Leu	Asp	Ser	Leu	Ala	Glu	Lys	Cys	Met	Gly
					500					505					510	
	Val	Leu	Tyr	Gln	Glu	His	His	Leu	Arg	Val	Tyr	Cys	Val	Met	Ile	Thr
5			515						520					525		
	Ala	Pro	Asn	Thr	Leu	Pro	Arg	Val	Leu	Lys	Asn	Gly	Arg	Gln	Glu	Ile
			530					535					540			
	Gly	Asn	Met	Leu	Cys	Arg	Lys	Glu	Phe	Asp	Asn	Gly	Ser	Leu	Pro	Cys
	545					550					555				560	
10	Glu	His	Val	Lys	Phe	Ser	Val	Glu	Arg	Ser	Val	Leu	Ser	Leu	Pro	Ile
					565					570					575	
	Gly	Val	Asp	Pro	Val	Gly	Gly	Ile	Trp	Ser	Val	Pro	Ser	Ser	Ala	Ala
				580					585					590		
	Arg	Gln	Asp	Ala	Leu	Ala	Met	Gln	Glu	Lys	Gln	Tyr	Ser	Gly	Val	Asp
15			595					600						605		
	Leu	Arg	Asp	Val	Ile	Met	Asp	Asp	Arg	Thr	Ser	Thr	Pro	Leu	Asn	Asn
			610					615					620			
	Phe	Asn	Ser	Ile	Val	Asp	Leu	Leu	Gln	Trp	Arg	Val	Ser	Arg	Gln	Gly
	625					630					635				640	
20	Glu	Glu	Leu	Cys	Tyr	Cys	Ser	Ile	Asp	Gly	Arg	Gly	Arg	Glu	Gly	Lys
					645						650				655	
	Gly	Ile	Thr	Trp	Lys	Lys	Phe	Asp	Ser	Lys	Val	Ala	Ala	Val	Ala	Ala
				660					665					670		
	Tyr	Leu	Lys	Asn	Lys	Val	Lys	Leu	Arg	Pro	Gly	Asp	His	Val	Ile	Leu
25			675					680						685		
	Met	Tyr	Thr	His	Ser	Glu	Glu	Tyr	Val	Phe	Ala	Val	His	Ala	Cys	Phe
			690					695					700			
	Cys	Leu	Gly	Leu	Val	Ala	Ile	Pro	Ile	Ser	Pro	Val	Asp	Gln	Asn	Arg
	705					710					715				720	
30	Leu	Ser	Glu	Asp	Ala	Pro	Ala	Leu	Leu	His	Val	Ile	Val	Asp	Phe	Arg
					725						730				735	
	Val	Lys	Ala	Ile	Leu	Val	Asn	Gly	Glu	Val	Asn	Asp	Leu	Leu	Lys	Gln
				740					745					750		
	Lys	Ile	Val	Ser	Gln	His	Ile	Lys	Gln	Ser	Ala	His	Val	Val	Arg	Thr
35			755					760						765		
	Ser	Val	Pro	Ser	Val	Tyr	Asn	Thr	Ser	Lys	Pro	Pro	Lys	Gln	Ser	His
			770					775					780			
	Gly	Cys	Arg	His	Leu	Gly	Phe	Thr	Met	Asn	Pro	Gln	Trp	Leu	Asn	Ser
	785					790					795				800	
40	Lys	Gln	Pro	Ala	Val	Ile	Trp	Thr	Tyr	Trp	Thr	Pro				

40

805

810

<210> 48

<211> 1836

5<212> DNA

<213> *Cochliobolus heterostrophus*

<400> 48

```

atgtctctct cgggcctgct gcgctcgcg gaggcacccg ctgccaagcg tcacctcctc      60
10tccaactgga atgccgccca gtttgaggag ctcaagtact cgtacggcct cactgggtgct      120
gaccaagtgc gcaacttctt gtgggtcgac acctttctct acatgctcat tggcatctct      180
ggcatgctcc tcatgctccg catctccaac atggtctgga agcacagccg gcacatcacc      240
gcaatgggaa gcccaaggca aaagtactgg gagaccaacc gaacaagctg gtggccctgg      300
ctcaaccgcc acatcctcgt cgccccgctc tggagaaga agcacaacgc ccagttccag      360
15atcagcagcg cgattgacaa cggaaccctc cctggaagat ggcacaccat catgctcctc      420
atctacgtcg gcctcaacgt tgcattggtg cttgccctcc cctacgacgt cctcgaccac      480
agggagacgc tcgccgccct tcgtggacgc tctggaaccc tcgccgccct caacctcatc      540
cccaccatcc tcttcgccct ccgcaacaac cccctcatct cccttctcca ggtctcgtac      600
gacgacttca accttttcca ccgctgggct gcccgaatca ccattgccga ggccattgtc      660
20cacactgccg cttggttgta caacaccaag gctggcggtg gatggcacgc cgtcgtagct      720
gccctccaca ccgagggctc ttacggatgg ggcattggcg gaactgtcgc cttcaccttc      780
atcggcatcc aggcctggtc cccattccgt cagcctttt acgagacctt tctcaacatc      840
caccgcgtca tggtcattgc tgctctcctc ggcttgtaga agcacctgga gctgcacgct      900
ctgccccagg tcccatggat gtacctcatc ttcattctct gggcggtga gtggttcctc      960
25cgctgtgct ccatctgcta ctacggcttc agcctgaagc aacgctcttc catcacgctc      1020
gaggccttgc ctggcgaagc tgtccgtcta accatcaaca tggtcgcga atggaccccc      1080
cgtcccggat gtcacgtgca catgtggatg cctcgccctc ccctctggtc ctcgcatcca      1140
ttttccgtcg cctgggctgc gacctgacc gacgactcca aagagatgac gcttcccact      1200
ctggaaggcg acgtcaccat gatcaatggc caaccagga aatcaaaaca aatcagtctc      1260
30atctgccgtg ccgtaccgg actcaccctg caaatgtatg aaaaggcaag caaaagcccc      1320
aacgagcaat tcaccacatg gggcttcatt gaaggcccat acggtggtca ccacagtctt      1380
gactcgtagc gtacttgtgt actgtttgcc gcagggtgtag gcatcaccca ccaggatcatg      1440
tacctcaagc atctagtcaa tggcttcaac aacggcacca ctgccacgca aaagattgtc      1500
ctcatctgga cagtaccac gcccgactgc ctggagtggg tgcgcccatg gatggacgaa      1560
35gtcctccgca tgaagggtcg caagcagtgt ctccgcatca agctcttcat ctccagacca      1620
aagggccgtg tcgagagcag tagcgacact gtcaagatgt acagcggcag gcccaacatg      1680
aggagcttgt tggaggagga ggccaagcac cgcgttgggt ccatggccgt gaccgtgtgc      1740
gcgtctggcg gcatggccga cgggtgtacg catgcagtgc gccactgct taccgagggt      1800
tcggttgatt tcatagagga agcctttacg tattga      1836

```

40

41

<210> 49

<211> 611

<212> PRT

<213> Cochliobolus heterostrophus

5

<400> 49

```

Met Ser Leu Ser Gly Leu Leu Arg Ser Arg Glu Ala Pro Ala Ala Lys
 1           5           10           15
Arg His Leu Leu Ser Asn Trp Asn Ala Ala Gln Phe Glu Glu Leu Lys
10           20           25           30
Tyr Ser Tyr Gly Leu Thr Gly Val Asp Gln Val Gly Asn Phe Leu Trp
      35           40           45
Val Asp Thr Phe Leu Tyr Met Leu Ile Gly Ile Ser Gly Met Leu Leu
      50           55           60
15Met Leu Arg Ile Ser Asn Met Val Trp Lys His Ser Arg His Ile Thr
   65           70           75           80
Ala Met Gly Ser Pro Arg Gln Lys Tyr Trp Glu Thr Asn Arg Thr Ser
      85           90           95
Trp Trp Pro Trp Leu Asn Arg His Ile Leu Val Ala Pro Leu Trp Lys
20           100          105          110
Lys Lys His Asn Ala Gln Phe Gln Ile Ser Ser Ala Ile Asp Asn Gly
      115          120          125
Thr Leu Pro Gly Arg Trp His Thr Ile Met Leu Leu Ile Tyr Val Gly
      130          135          140
25Leu Asn Val Ala Trp Cys Leu Ala Leu Pro Tyr Asp Val Leu Asp His
   145          150          155          160
Arg Glu Thr Leu Ala Ala Leu Arg Gly Arg Ser Gly Thr Leu Ala Ala
      165          170          175
Leu Asn Leu Ile Pro Thr Ile Leu Phe Ala Leu Arg Asn Asn Pro Leu
30           180          185          190
Ile Ser Leu Leu Gln Val Ser Tyr Asp Asp Phe Asn Leu Phe His Arg
      195          200          205
Trp Ala Ala Arg Ile Thr Ile Ala Glu Ala Ile Val His Thr Ala Ala
      210          215          220
35Trp Leu Tyr Asn Thr Lys Ala Gly Gly Gly Trp His Ala Val Val Ala
   225          230          235          240
Ala Leu His Thr Glu Gly Ser Tyr Gly Trp Gly Met Gly Gly Thr Val
      245          250          255
Ala Phe Thr Phe Ile Gly Ile Gln Ala Trp Ser Pro Phe Arg His Ala
40           260          265          270

```

42

Phe Tyr Glu Thr Phe Leu Asn Ile His Arg Val Met Val Ile Ala Ala
 275 280 285
 Leu Leu Gly Leu Tyr Lys His Leu Glu Leu His Ala Leu Pro Gln Val
 290 295 300
 5Pro Trp Met Tyr Leu Ile Phe Ile Phe Trp Ala Ala Glu Trp Phe Leu
 305 310 315 320
 Arg Leu Cys Ser Ile Cys Tyr Tyr Gly Phe Ser Leu Lys Gln Arg Ser
 325 330 335
 Ser Ile Thr Val Glu Ala Leu Pro Gly Glu Ala Val Arg Leu Thr Ile
 10 340 345 350
 Asn Met Val Arg Glu Trp Thr Pro Arg Pro Gly Cys His Val His Met
 355 360 365
 Trp Met Pro Arg Leu Ser Leu Trp Ser Ser His Pro Phe Ser Val Ala
 370 375 380
 15Trp Ala Ala Thr Leu Thr Asp Asp Ser Lys Glu Met Thr Leu Pro Thr
 385 390 395 400
 Leu Glu Gly Asp Val Thr Met Ile Asn Gly Gln Pro Arg Lys Ser Lys
 405 410 415
 Gln Ile Ser Leu Ile Cys Arg Ala Arg Thr Gly Leu Thr Arg Gln Met
 20 420 425 430
 Tyr Glu Lys Ala Ser Lys Ser Pro Asn Glu Gln Phe Thr Thr Trp Gly
 435 440 445
 Phe Ile Glu Gly Pro Tyr Gly Gly His His Ser Leu Asp Ser Tyr Gly
 450 455 460
 25Thr Cys Val Leu Phe Ala Ala Gly Val Gly Ile Thr His Gln Val Met
 465 470 475 480
 Tyr Leu Lys His Leu Val Asn Gly Phe Asn Asn Gly Thr Thr Ala Thr
 485 490 495
 Gln Lys Ile Val Leu Ile Trp Thr Val Pro Thr Pro Asp Cys Leu Glu
 30 500 505 510
 Trp Val Arg Pro Trp Met Asp Glu Val Leu Arg Met Lys Gly Arg Lys
 515 520 525
 Gln Cys Leu Arg Ile Lys Leu Phe Ile Ser Arg Pro Lys Gly Arg Val
 530 535 540
 35Glu Ser Ser Ser Asp Thr Val Lys Met Tyr Ser Gly Arg Pro Asn Met
 545 550 555 560
 Arg Ser Leu Leu Glu Glu Glu Ala Lys His Arg Val Gly Ala Met Ala
 565 570 575
 Val Thr Val Cys Ala Ser Gly Gly Met Ala Asp Gly Val Arg His Ala
 40 580 585 590

Val Arg Pro Leu Leu Thr Glu Gly Ser Val Asp Phe Ile Glu Glu Ala
 595 600 605
 Phe Thr Tyr
 610

5
 <210> 50
 <211> 6553
 <212> DNA
 <213> Cochliobolus heterostrophus

10
 <400> 50

tgcctgcgcc	tgtgcttgtg	cctgtggaat	gtcgcggccc	gctgctgcat	agcctatctg	60
tacatacaac	accatcccat	cccgttcac	ctgccttgcc	tccctcctcg	tgccacacat	120
ccgccgccc	caacaccatg	gctgcgacca	accccgagct	gcaggccaaa	ctgcaggagc	180
15tggaaccacga	gctcgaggag	ggcgatatta	cacaaaaagg	gtccgtactg	ctgcaccacc	240
accgccatcc	gcctctctgc	gtgcgcta	cagtcgcata	gctatgaaaa	acgtcgcacc	300
gtgctgctgt	cgcagtatct	agggcctgac	tttgetgccc	agttgcaggc	cgacctgaac	360
cagcagaacc	caccccaacc	atccagttag	ggctctcgct	cccgccaccg	atcctttgct	420
attccgtccg	gtccgagtcc	atcacggcga	ccacaacccc	cacatatcca	gctccccgcg	480
20cccgactcat	accatgacgc	ttccgcacag	ggccaattgg	gcgcacccat	gccatagcg	540
aacgcctccg	ccgctgcctc	ggggggctcg	cagtacatgg	catacccgcc	cagccaagtc	600
ggccggttttc	aagagaagca	gctgggcctg	cgtacaaatt	cgctccagcg	caattcctca	660
cagctgtcgc	aaggaagcga	gacgttcatt	ccacggcctc	aaacgcctga	atacaaccac	720
tcgcgcgagc	ccaccatgat	gggcaactac	gccttcaatc	cagacaatca	gcaaagttat	780
25gatggccaat	ttggctctcc	gggagaggcc	agtcgaagga	gcaccatgct	cgaggtaa	840
cagggttatt	tttccgactt	cacaggccag	cagatgcaag	acaatcgcg	ctcgtatggg	900
ggacccaacc	gctactcgtc	gggagatgcc	ttttctccta	ccgccgcgat	tccacctccc	960
atgatgaacc	ccaacgatct	ccccttgggc	gctgctgaaa	ccatgatgcc	gctagagccc	1020
cgcgatctgc	cttttgacgt	ttacgacctt	cacaacccca	atgtcaaaat	gtcaaagttt	1080
30gacaacattg	gcgctgtctt	gcgtcaccga	agtcgcacac	agccaaggac	gactgccttc	1140
tgggtccttg	acgcaaaaagg	caaagagacg	gcgtccatca	cctgggaaaa	ggtggctagt	1200
cgcgcggaaa	aggtggccaa	agtgattcgg	gacaagagca	acctctatcg	aggcgaccgt	1260
gtggcattag	tgtacaggga	tacagaaatc	attgattttg	tcgtggcggt	gatgggctgc	1320
ttcattgcgg	gcgttgtagc	ggtaccatc	aatagcgctg	acgactacca	gaaactcatt	1380
35cttctcctaa	cgacaactca	agctcatctc	gcattgacca	cagacaacaa	tctcaaggcc	1440
tttcatcgctg	acattagtca	gaaccgtctg	aaatggccga	gtggggtaga	gtgggtggaag	1500
acgaacgagt	ttggcagcca	ccaccccaag	aaacatgacg	atactccagc	tttgcaagta	1560
ccagagggttg	cctatattga	gttctcgcgt	gcacctactg	gtgaccttcg	cggtgtgggtg	1620
cttagtcacc	ggactattat	gcaccaa	atg	gtgccatgat	tagcacgata	1680
40ccacccaacg	ctcagagcca	agacacgttc	agcactagcc	tacgggatgc	agaggggaaag	1740

	ttcgttgctc	cagcaccgtc	cagaaacccc	acagaagtga	tcctcacgta	cctcgacccg	1800
	cgcgaaagcg	ctggtctcat	tctcagtgtc	ttgtttgcag	tttatggagg	ccacaccacc	1860
	gtatggctcg	agacagcgac	catggaaacc	ccgggtctat	atgcacatct	catcaccaaa	1920
	tacaagtcca	acatactgct	agcggattac	ccaggcctca	agcgcgctgc	atacaactac	1980
5	caacaggatc	caatggctac	aagaaacttc	aagaaaaaca	cagaacccaa	cttcgcctcc	2040
	gtgaagatct	gtctgattga	cacgcttacc	gtcgactgtg	aattttcacga	aattctcgga	2100
	gatcgatatt	tcaggccact	gcgaaaccct	agagcgcgag	aactgatcgc	gccaatgctc	2160
	tgcttgccag	aacatgggtg	aatgataata	tctgtacgcg	actggctagg	tgagaggag	2220
	cgcattgggt	gcccgcctaag	catagcagta	gaagagtcag	ataatgatga	agatgataca	2280
10	gaggataagt	atgcagcggc	aaatggctac	tccagtctta	ttggtggtgg	cactacaaag	2340
	aacaaaaagg	agaagaagaa	gaaaggcccc	acagagctta	cagaaatctt	gctggacaag	2400
	gaagctctga	agatgaacga	agtcattgtt	ctggccattg	gagaagaagc	aagcaagcgg	2460
	gcaaacgagc	ccggcaccat	gcgagtcggt	gcctttggat	accccatacc	ggatgcgaca	2520
	ctagctattg	tagaccctga	gacaagtctt	ctatgttcac	catactcgat	aggcgagatc	2580
15	tgggtagatt	cgccttcact	ctctggtggc	ttctggcagc	tgcagaagca	tacagagacc	2640
	attttccatg	ctcgaccata	ccgtttcgtt	gagggtagcc	ctacgccaca	gttgcttgaa	2700
	ctcgagtttc	tgcgtactgg	actcctcggc	tttgttgtag	agggaaaaat	atltgtcctt	2760
	ggactgtacg	aagatcgcat	cagacagcgt	gttgaatggg	tagaaaatgg	tcagcttgaa	2820
	gccgagcatt	gatacttttt	tgtgcagcac	ctggtcacao	gcattatgaa	ggccgtgcca	2880
20	aaaattttacg	actggtaagt	gagctgcca	cagagcaagg	actgtctaac	gtgtcatagc	2940
	tcgtcgtttg	attcttatgt	aaatggtgaa	tacctgccaa	tcattctcat	cgagacgcag	3000
	gccgcacoga	ctgcgccac	aaaccaggt	ggaccaccac	aacaattgga	tataccattt	3060
	ttggattcac	tatctgagag	gtgcatggag	gtcctttacc	aagagcatca	tttacgggta	3120
	tactgcgtga	tgattacagc	acctaataca	cttcacagag	tcatacaaga	cggacggcga	3180
25	gaaattggca	atatgctgtg	taggagagag	tttgacaatg	gctctctgcc	ctgtgtacac	3240
	gtaaagtttg	gcattgagcg	atcagtgcag	aacattgccc	tcggtgacga	tcctgctggc	3300
	ggcatgtggg	catttgaggc	atcaatggca	cgtcagcaat	tcttgatgct	ccaagacaag	3360
	caatactctg	gtgtcgatca	tcgcgaagtc	gtcattgacg	acaggacatc	gactccactc	3420
	aatcagttct	cgaatatcca	cgacctgatg	caatggcggt	tatctcggca	ggccgaggaa	3480
30	cttgcttact	gcactgtcga	cggtcgagga	aaagagggca	aaggcgtcaa	ttggaagaag	3540
	tttgatcaaa	aggttgcggg	cgtagcaatg	tacctcaaga	acaagggtcaa	ggtccaggcc	3600
	ggcgatcatc	tccttctgat	gtacacgcat	tcagaagaat	ttgtttatgc	tgttcatgca	3660
	tgttttgtgc	ttggagctgt	ttgcatacca	atggcgccaa	ttgatcagaa	ccggttgaat	3720
	gaggatgcgc	cggccttgct	gcataatcct	gcagatttca	agggtcaaagc	cattcttgct	3780
35	aacgctgacg	ttgaccatct	gatgaagatc	aagcaagtat	cgcagcacat	caaacaatcg	3840
	gccgctatcc	tcaagatcag	tgtgccaaac	acatacagca	caacaaagcc	gccaaagcaa	3900
	tccagtggct	gccgcgacct	caagcttaca	attcgaccgg	catggattca	ggcgggtttc	3960
	ccagtgctag	tctggacata	ctggacgccc	gatcaacgtc	gtatcgcagt	tcagctgggc	4020
	catagccaaa	tcattggcact	gtgcaaggtc	caaaaagaaa	catgccaaat	gacaagtaca	4080
40	cgaccagttc	ttggttgtgt	ccggagcacg	ataggacttg	gtttccttca	catttgtctc	4140

atgggaatct	tccttgccgc	acccacatac	ctggtgtcac	ctgttgactt	tgcacaaaac	4200
cctaatatct	tgttccaaac	gcttttcgcg	tacaagatca	aggatgcata	tgcaacgagt	4260
caaatgttgg	accacgccat	cgcacgcgga	gctggtaaga	gtatggctct	gcacgagctg	4320
aagaatctca	tgattgcgac	tgatggaaga	ccacgcgttg	atgtttgtaa	gtgaacattt	4380
5gtatgagagg	actttcatga	ttgctaactc	aatgcagacc	aaagagtgcg	tgtgcacttt	4440
gcgccagcca	acttagaccc	aaccgcaatc	aacactgtct	actcacatgt	attgaaccca	4500
atggtagcat	cacgatcata	catgtgtatt	gagccagtcg	agctccatct	cgatgtgcat	4560
gctctgcgac	gcggcctcgt	catgcccggt	gaccctgaca	cagagcccaa	cgctttgctc	4620
gtccaagact	cgggcatggg	gccagtgcgc	acgcaaatat	ccattgtcaa	cccagagacc	4680
10aaccaactgt	gcttgaacgg	cgagtacggc	gagatctggg	tgcagtccga	ggcgaatgct	4740
tatagcttct	acatgtcgaa	agagcgcttg	gatgcagaac	gcttcaatgg	gaggacgatt	4800
gacggagacc	caaatgtgcg	atatgttcgt	acaggcgatt	taggattttt	gcacagcgtg	4860
acacggccca	ttggacccaa	cggtgcacct	gttgatatgc	aggtgctttt	cgtgcttgga	4920
agcatagggtg	acacttttga	agtcaacgga	ctgaaccatt	tctctatgga	cattgagcag	4980
15tctgttgaac	gttgtcaccg	gaatattgtc	cctggagggt	ggtacgtttc	ttcgattcgc	5040
tgttatttag	taaatactta	ctaactctct	acagtgtctg	tttccaggca	ggtgggcttg	5100
ttgttgcgt	tgtggaaatc	ttccgacgca	acttctctgc	aagcatgggtg	cctgtgattg	5160
tcaatgcaat	tttgaacgag	catcagctgg	tcattgacat	tgtctcgttt	gtgcaaaagg	5220
gcgacttcca	cgggtctcgt	ctgggcgaga	agcaacgcgg	aaagattctt	gcaggatggg	5280
20tcacacggaa	gatgcgcaca	atagcccagt	acagtatacg	ggatccctaat	ggacaggatt	5340
cccagatgat	gatcacggaa	gagcctgggtc	cacgggctag	catgactgga	agtatgcttg	5400
ggcgaatggg	cggcccagcc	agtatcaagg	cggggtcgac	aagagcaccg	agtctaattg	5460
gcatgacagc	gactatgaat	aatctatccc	ttacacagca	gcaacagcag	caataccaac	5520
agccgggtat	gtatgtctca	cagcaaggca	tgcaccccca	gcaacaacac	caatttagca	5580
25tgtccaacac	gccaccacaa	ggtccacccc	aaggcgtaga	actacatgat	cctagcgacc	5640
gcacaccaac	agacaaccgg	cactcttttc	ttgcccagccc	gcgtatgcag	aaccagggcc	5700
aaatgaacga	gacggggcgc	tacgaaccca	tgaactatca	aaacgcgtat	catccgcctc	5760
aacaacaata	cgaatctgaa	gacgggggga	gcagactcag	cggccccgtg	ccagacgtgc	5820
tgcgggccggg	tccttcatcc	gggtccatag	agcagcacga	ccaagctaac	aacgacaaca	5880
30atatgtggaa	taatcgcgag	tactatggta	acagcccatc	gtatgcaggc	ggatacacgc	5940
aagatggcaa	tatccacgag	cagcaacaac	acgatgagta	cacgagtaat	gcgtcatatg	6000
gcggaaatca	aggagcaggc	ggaggcagcg	gcggcggttg	cgggtctccga	gttgcaaatc	6060
gtgacagctc	cgacagcgag	ggtgcagatg	acgacgcttg	gagacgtgat	gcccttgctc	6120
agatcaattt	tgcgggcggc	gctgctgctg	cctccgctgg	agcacctgct	gctggtgctt	6180
35cttcttcgca	gccgggccat	gcgcagtaga	cgggatatgc	gtgagttttt	ttttaaattt	6240
cgtacataga	gaccgttgta	tacgcagggt	tcaaattaga	agagcgaata	tgcatatcag	6300
ctgttggttca	atgttctagt	ttgggaagggt	taaccccccc	cccttcccct	tccaagactt	6360
ttcacttggt	tgtgtgtgat	ttaaatctgg	agatttcaaa	tctacatctc	gctatacata	6420
ggtgttggtt	gataacgtag	ggggcagaag	ggtatctcgt	gatattagac	tgggagttgc	6480
40atgaatcaag	gtgttgagca	aaaaaagaga	gagcggtgaa	gggcgggggg	gatagggtgt	6540

gtgcacgtgg ctg

6553

5<210> 51

<211> 530

<212> PRT

<213> Alternaria solani

10<400> 51

	Lys	Lys	Lys	Gly	Pro	Thr	Glu	Leu	Thr	Glu	Ile	Leu	Leu	Asp	Lys	Glu
	1				5					10					15	
	Ala	Leu	Lys	Leu	Asn	Glu	Val	Val	Val	Leu	Ala	Ile	Gly	Glu	Glu	Val
				20					25					30		
15	Ser	Lys	Arg	Val	Asn	Glu	Pro	Gly	Thr	Met	Arg	Val	Gly	Ala	Phe	Gly
			35					40					45			
	Tyr	Pro	Ile	Pro	Asp	Ala	Thr	Leu	Ala	Val	Val	Asp	Pro	Glu	Thr	Asn
		50					55					60				
	Leu	Leu	Cys	Ser	Pro	Tyr	Ser	Ile	Gly	Glu	Ile	Trp	Val	Asp	Ser	Pro
20	65					70					75				80	
	Ser	Leu	Ser	Gly	Gly	Phe	Trp	Gln	Leu	Gln	Lys	His	Thr	Glu	Thr	Ile
				85					90					95		
	Phe	His	Ala	Arg	Pro	Tyr	Arg	Phe	Val	Glu	Gly	Ser	Pro	Thr	Pro	Gln
				100					105					110		
25	Leu	Leu	Glu	Leu	Glu	Phe	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Cys	Val	Val
			115					120					125			
	Glu	Gly	Lys	Ile	Phe	Val	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Ile	Arg	Gln
		130					135					140				
	Arg	Val	Glu	Trp	Val	Glu	His	Gly	Gln	Leu	Glu	Ala	Glu	His	Arg	Tyr
30	145					150					155				160	
	Phe	Phe	Val	Gln	His	Leu	Val	Thr	Ser	Ile	Met	Lys	Ala	Val	Pro	Lys
				165						170					175	
	Ile	Tyr	Asp	Cys	Ser	Ser	Phe	Asp	Ser	Tyr	Val	Asn	Gly	Glu	Tyr	Leu
				180					185					190		
35	Pro	Ile	Ile	Leu	Ile	Glu	Thr	Gln	Ala	Ala	Ser	Thr	Ala	Pro	Thr	Asn
			195					200					205			
	Pro	Gly	Gly	Pro	Pro	Gln	Gln	Leu	Asp	Ile	Pro	Phe	Leu	Asp	Ser	Leu
		210					215					220				
	Ser	Glu	Arg	Cys	Met	Glu	Val	Leu	Tyr	Gln	Glu	His	His	Leu	Arg	Val
40	225					230					235					240

47

Tyr Cys Val Met Ile Thr Ala Pro Asn Thr Leu Pro Arg Val Ile Lys
 245 250 255
 Asn Gly Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe Asp
 260 265 270
 5Asn Gly Ser Leu Pro Cys Val His Val Lys Phe Gly Val Glu Arg Ser
 275 280 285
 Val Gln Asn Ile Ala Leu Gly Asp Asp Pro Ala Gly Gly Met Trp Ser
 290 295 300
 Tyr Glu Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys
 10305 310 315 320
 Gln Tyr Ser Gly Val Asp His Arg Glu Val Val Ile Asp Asp Arg Thr
 325 330 335
 Ser Thr Pro Leu Asn Gln Phe Ser Asn Ile His Asp Leu Met Gln Trp
 340 345 350
 15Arg Val Gln Arg Gln Ala Glu Glu Leu Ala Tyr Cys Thr Val Asp Gly
 355 360 365
 Arg Gly Lys Glu Gly Lys Gly Val Asn Trp Lys Lys Phe Asp Gln Lys
 370 375 380
 Val Ala Gly Val Ala Met Tyr Leu Lys Asn Lys Val Lys Gly Gln Thr
 20385 390 395 400
 Gly Asp His Leu Leu Leu Met Tyr Thr His Ser Glu Asp Phe Val Tyr
 405 410 415
 Ala Val His Ala Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala
 420 425 430
 25Pro Ile Asp Gln Asn Arg Leu Asn Glu Asp Ala Pro Ala Leu Leu His
 435 440 445
 Ile Ile Ala Asp Phe Lys Val Lys Ala Ile Leu Val Asn Ala Gly Val
 450 455 460
 Asp His Leu Met Lys Val Lys Gln Val Ser Gln His Ile Lys Gln Ser
 30465 470 475 480
 Ala Val Ile Leu Lys Ile Asn Val Pro Asn Thr Tyr Asn Thr Thr Lys
 485 490 495
 Pro Pro Lys Gln Ser Ser Gly Cys Arg Asp Leu Lys Leu Thr Ile Arg
 500 505 510
 35Pro Ala Trp Ile Gln Ser Gly Phe Pro Val Leu Val Trp Thr Tyr Trp
 515 520 525
 Thr Pro
 530

<211> 530

<212> PRT

<213> Pyrenophora teres

5<400> 52

Lys	Lys	Lys	Gly	Pro	Thr	Glu	Leu	Thr	Glu	Ile	Leu	Leu	Asp	Lys	Glu
1				5					10					15	
Ala	Leu	Lys	Met	Asn	Asp	Val	Val	Val	Leu	Ala	Ile	Gly	Glu	Glu	Ala
			20					25					30		
10Ser	Lys	Arg	Ala	Asn	Glu	Pro	Gly	Thr	Met	Arg	Val	Gly	Ala	Phe	Gly
			35				40					45			
Tyr	Pro	Ile	Pro	Asp	Ala	Thr	Leu	Ala	Val	Val	Asp	Pro	Glu	Thr	Asn
	50					55					60				
Leu	Leu	Cys	Ser	Pro	Tyr	Ser	Ile	Gly	Glu	Ile	Trp	Val	Asp	Ser	Pro
1565					70					75				80	
Ser	Leu	Ser	Gly	Gly	Phe	Trp	Gln	Leu	Gln	Lys	His	Thr	Glu	Thr	Ile
				85					90				95		
Phe	His	Ala	Arg	Pro	Tyr	Arg	Phe	Val	Glu	Gly	Ser	Pro	Thr	Pro	Gln
			100					105					110		
20Leu	Leu	Glu	Leu	Glu	Phe	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Phe	Val	Val
		115					120					125			
Glu	Gly	Lys	Val	Phe	Ile	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Ile	Arg	Gln
	130					135					140				
Arg	Val	Glu	Trp	Val	Glu	His	Gly	Gln	Leu	Glu	Ala	Glu	His	Arg	Tyr
25145					150					155				160	
Phe	Phe	Val	Gln	His	Leu	Val	Thr	Ser	Ile	Met	Lys	Ala	Val	Pro	Lys
			165						170				175		
Ile	Tyr	Asp	Cys	Ser	Ser	Phe	Asp	Ser	Tyr	Val	Asn	Gly	Glu	Tyr	Leu
		180						185				190			
30Pro	Ile	Ile	Leu	Ile	Glu	Thr	Gln	Ala	Ala	Ser	Thr	Ala	Pro	Thr	Asn
		195					200					205			
Pro	Gly	Gly	Pro	Pro	Gln	Gln	Leu	Asp	Ile	Pro	Phe	Leu	Asp	Ser	Leu
	210					215					220				
Ser	Glu	Arg	Cys	Met	Glu	Val	Leu	Tyr	Gln	Glu	His	His	Leu	Arg	Val
35225					230					235				240	
Tyr	Cys	Val	Met	Ile	Thr	Ala	Pro	Asn	Thr	Leu	Pro	Arg	Val	Val	Lys
			245						250				255		
Asn	Gly	Arg	Arg	Glu	Ile	Gly	Asn	Met	Leu	Cys	Arg	Arg	Glu	Phe	Asp
		260					265					270			
40Asn	Gly	Ser	Leu	Pro	Cys	Val	His	Val	Lys	Phe	Gly	Val	Glu	Arg	Ser

49

275	280	285
Val Leu Asn Ile Ala Leu Gly Asp Asp Pro Ser Gly Gly Met Trp Ser		
290	295	300
Tyr Glu Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys		
5305	310	315
Gln Tyr Ser Gly Val Asp His Arg Glu Val Val Met Asp Asp Arg Thr		
325	330	335
Ser Thr Pro Leu Asn Gln Phe Ser Asn Ile His Asp Leu Met Gln Trp		
340	345	350
10Arg Val Ser Arg Gln Ala Glu Glu Leu Ala Tyr Cys Thr Val Asp Gly		
355	360	365
Arg Gly Lys Glu Gly Lys Gly Val Asn Trp Lys Lys Phe Asp Gln Lys		
370	375	380
Val Ala Gly Val Ala Met Tyr Leu Lys Asn Lys Val Lys Val Gln Thr		
15385	390	395
Gly Asp His Leu Leu Leu Met Tyr Thr His Ser Glu Asp Phe Val Tyr		
405	410	415
Ala Val His Ala Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala		
420	425	430
20Pro Ile Asp Gln Asn Arg Leu Asn Glu Asp Ala Pro Ala Leu Leu His		
435	440	445
Ile Leu Ala Asp Phe Lys Val Lys Ala Ile Leu Val Asn Ala Asp Val		
450	455	460
Asp His Leu Met Lys Val Lys Gln Val Ser Gln His Ile Lys Gln Ser		
25465	470	475
Ala Ala Ile Phe Lys Ile Asn Val Pro His Thr Tyr Asn Thr Thr Lys		
485	490	495
Pro Pro Lys Gln Ser Ser Gly Cys Arg Asp Leu Lys Leu Thr Ile Arg		
500	505	510
30Pro Ala Trp Val Gln Pro Gly Phe Pro Val Leu Val Trp Thr Tyr Trp		
515	520	525
Thr Pro		
530		

35<210> 53

<211> 531

<212> PRT

<213> Fusarium graminearum

40<400> 53

50

Glu Glu Arg Ala Lys Asn Glu Leu Gly Glu Val Leu Leu Asp Arg Glu
 1 5 10 15
 Ala Leu Lys Thr Asn Glu Val Val Val Val Ala Ile Gly Asn Asp Ala
 20 25 30
 5Arg Lys Arg Val Thr Asp Asp Pro Gly Leu Val Arg Val Gly Ser Phe
 35 40 45
 Gly Tyr Pro Ile Pro Asp Ala Thr Leu Ser Val Val Asp Pro Glu Thr
 50 55 60
 Gly Leu Leu Ala Ser Pro His Ser Val Gly Glu Ile Trp Val Asp Ser
 1065 70 75 80
 Pro Ser Leu Ser Gly Gly Phe Trp Ala Gln Pro Lys Asn Thr Glu Leu
 85 90 95
 Ile Phe His Ala Arg Pro Tyr Lys Phe Asp Pro Gly Asp Pro Thr Pro
 100 105 110
 15Gln Pro Val Glu Pro Glu Phe Leu Arg Thr Gly Leu Leu Gly Thr Val
 115 120 125
 Ile Glu Gly Lys Ile Phe Val Leu Gly Leu Tyr Glu Asp Arg Ile Arg
 130 135 140
 Gln Lys Val Glu Trp Val Glu His Gly His Glu Leu Ala Glu Tyr Arg
 20145 150 155 160
 Tyr Phe Phe Val Gln His Ile Val Val Ser Ile Val Lys Asn Val Pro
 165 170 175
 Lys Ile Tyr Asp Cys Ser Ala Phe Asp Val Phe Val Asn Asp Glu His
 180 185 190
 25Leu Pro Val Val Val Leu Glu Ser Ala Ala Ala Ser Thr Ala Pro Leu
 195 200 205
 Thr Ser Gly Gly Pro Pro Arg Gln Pro Asp Thr Ala Leu Leu Glu Ser
 210 215 220
 Leu Ala Glu Arg Cys Met Glu Val Leu Met Ser Glu His His Leu Arg
 30225 230 235 240
 Leu Tyr Cys Val Met Ile Thr Ala Pro Asp Thr Leu Pro Arg Val Val
 245 250 255
 Lys Asn Gly Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe
 260 265 270
 35Asp Leu Gly Asn Leu Pro Cys Val His Val Lys Phe Gly Val Glu His
 275 280 285
 Ala Val Leu Asn Leu Pro Ile Gly Val Asp Pro Ile Gly Gly Ile Trp
 290 295 300
 Ser Pro Leu Ala Ser Asp Ser Arg Ala Glu Phe Leu Leu Pro Ala Asp
 40305 310 315 320

51

Lys	Gln	Tyr	Ser	Gly	Val	Asp	Arg	Arg	Glu	Val	Val	Ile	Asp	Asp	Arg	
				325				330				335				
Thr	Ser	Thr	Pro	Leu	Asn	Asn	Phe	Ser	Cys	Ile	Ser	Asp	Leu	Ile	Gln	
				340				345				350				
5Trp	Arg	Val	Ala	Arg	Gln	Pro	Glu	Glu	Leu	Ala	Tyr	Cys	Thr	Ile	Asp	
				355				360				365				
Gly	Lys	Ser	Arg	Glu	Gly	Lys	Gly	Val	Thr	Trp	Lys	Lys	Phe	Asp	Thr	
				370				375				380				
Lys	Val	Ala	Ser	Val	Ala	Met	Tyr	Leu	Lys	Asn	Lys	Val	Lys	Val	Arg	
10385					390				395				400			
Pro	Gly	Asp	His	Ile	Ile	Leu	Met	Tyr	Thr	His	Ser	Glu	Glu	Phe	Val	
				405				410				415				
Phe	Ala	Ile	His	Ala	Cys	Ile	Ser	Leu	Gly	Ala	Ile	Val	Ile	Pro	Ile	
				420				425				430				
15Ala	Pro	Leu	Asp	Gln	Asn	Arg	Leu	Asn	Glu	Asp	Val	Pro	Ala	Phe	Leu	
				435				440				445				
His	Ile	Val	Ser	Asp	Tyr	Asn	Val	Lys	Ala	Val	Leu	Val	Asn	Ala	Glu	
				450				455				460				
Val	Asp	His	Leu	Ile	Lys	Val	Lys	Pro	Val	Ala	Ser	His	Ile	Lys	Gln	
20465					470				475				480			
Ser	Ala	Gln	Val	Leu	Lys	Ile	Thr	Ser	Pro	Ala	Ile	Tyr	Asn	Thr	Thr	
				485				490				495				
Lys	Pro	Pro	Lys	Gln	Ser	Ser	Gly	Leu	Arg	Asp	Leu	Arg	Phe	Thr	Ile	
				500				505				510				
25Asp	Pro	Ala	Trp	Ile	Arg	Pro	Gly	Tyr	Pro	Val	Ile	Val	Trp	Thr	Tyr	
				515				520				525				
Trp	Thr	Pro														
				530												

30<210> 54

<211> 531

<212> PRT

<213> Coccidioides immitis

35<400> 54

Lys Glu Gly Pro Arg Asn Asp Leu Gly Glu Val Leu Leu Asp Lys Glu
1 5 10 15
Ala Leu Lys Asn Asn Glu Ile Val Ile Leu Ala Ile Gly Glu Glu Ala
20 25 30
40Arg Arg Leu Ala Asp Thr Thr Pro Asn Ala Val Arg Val Gly Ala Phe

	Gly	Tyr	Pro	Ile	Pro	Asp	Ala	Thr	Leu	Ala	Ile	Val	Asp	Pro	Glu	Thr
	50						55					60				
	Gly	Leu	Leu	Cys	Thr	Pro	Asn	Val	Val	Gly	Glu	Ile	Trp	Val	Asp	Ser
565						70					75					80
	Pro	Ser	Leu	Ser	Gly	Gly	Phe	Trp	Ala	Leu	Pro	Lys	Gln	Thr	Glu	Ser
					85					90					95	
	Ile	Phe	His	Ala	Arg	Pro	Tyr	Arg	Phe	Gln	Gly	Gly	Gly	Pro	Thr	Pro
				100					105					110		
10	Val	Ile	Val	Glu	Pro	Glu	Phe	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Cys	Val
				115					120					125		
	Ile	Glu	Gly	Gln	Ile	Phe	Val	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Leu	Arg
				130					135				140			
	Gln	Lys	Val	Glu	Trp	Val	Glu	His	Gly	Val	Glu	Val	Ala	Glu	His	Arg
15	145					150					155					160
	Tyr	Phe	Phe	Val	Gln	His	Leu	Ile	Leu	Ser	Ile	Met	Lys	Asn	Val	Pro
					165						170				175	
	Lys	Ile	His	Asp	Cys	Ser	Ala	Phe	Asp	Val	Phe	Val	Asn	Glu	Glu	His
				180						185				190		
20	Leu	Pro	Val	Val	Val	Leu	Glu	Ser	Tyr	Thr	Ala	Ser	Thr	Ala	Pro	Val
				195					200					205		
	Ala	Ser	Gly	Gln	Ser	Pro	Arg	Gln	Leu	Asp	Val	Pro	Leu	Leu	Asp	Ser
				210					215				220			
	Leu	Ala	Glu	Lys	Cys	Met	Gly	Val	Leu	Tyr	Gln	Glu	His	His	Leu	Arg
25	225					230					235					240
	Val	Tyr	Cys	Val	Met	Ile	Thr	Ala	Pro	Asn	Thr	Leu	Pro	Arg	Val	Leu
					245						250				255	
	Lys	Asn	Gly	Arg	Gln	Glu	Ile	Gly	Asn	Met	Leu	Cys	Arg	Lys	Glu	Phe
				260						265				270		
30	Asp	Asn	Gly	Ser	Leu	Pro	Cys	Glu	His	Val	Lys	Phe	Ser	Val	Glu	Arg
				275					280					285		
	Ser	Val	Leu	Ser	Leu	Pro	Ile	Gly	Val	Asp	Pro	Val	Gly	Gly	Ile	Trp
				290					295				300			
	Ser	Val	Pro	Ser	Ser	Ala	Ala	Arg	Gln	Asp	Ala	Leu	Ala	Met	Gln	Glu
35	305					310					315					320
	Lys	Gln	Tyr	Ser	Gly	Val	Asp	Leu	Arg	Asp	Val	Ile	Met	Asp	Asp	Arg
					325						330				335	
	Thr	Ser	Thr	Pro	Leu	Asn	Asn	Phe	Asn	Ser	Ile	Val	Asp	Leu	Leu	Gln
				340						345				350		
40	Trp	Arg	Val	Ser	Arg	Gln	Gly	Glu	Glu	Leu	Cys	Tyr	Cys	Ser	Ile	Asp

53

355	360	365
Gly Arg Gly Arg Glu Gly Lys Gly Ile Thr Trp Lys Lys Phe Asp Ser		
370	375	380
Lys Val Ala Ala Val Ala Ala Tyr Leu Lys Asn Lys Val Lys Leu Arg		
5385	390	395
Pro Gly Asp His Val Ile Leu Met Tyr Thr His Ser Glu Glu Tyr Val		
	405	410
Phe Ala Val His Ala Cys Phe Cys Leu Gly Leu Val Ala Ile Pro Ile		415
	420	425
10Ser Pro Val Asp Gln Asn Arg Leu Ser Glu Asp Ala Pro Ala Leu Leu		430
	435	440
His Val Ile Val Asp Phe Arg Val Lys Ala Ile Leu Val Asn Gly Glu		445
	450	455
Val Asn Asp Leu Leu Lys Gln Lys Ile Val Ser Gln His Ile Lys Gln		460
15465	470	475
Ser Ala His Val Val Arg Thr Ser Val Pro Ser Val Tyr Asn Thr Ser		480
	485	490
Lys Pro Pro Lys Gln Ser His Gly Cys Arg His Leu Gly Phe Thr Met		495
	500	505
20Asn Pro Gln Trp Leu Asn Ser Lys Gln Pro Ala Val Ile Trp Thr Tyr		510
	515	520
Trp Thr Pro		525
	530	

25<210> 55

<211> 2073

<212> DNA

<213> Cochliobolus heterostrophus

30<400> 55

atacgtggtg gagccgtgca accgttgctg tgtgctgagt gctgagttgc ggtggagaat	60
gccccgtggg gtcgggatgg gtagcgctgc aggggttttag ctgagatgga ggggagagag	120
gggggggttg ggatgtttaa aaggatgggg aggggtgtgt tctgtgctt ggatgttacg	180
ctgttgcgct gcttacttgc tacgttgctc gtggcagccg actcagtctt tctacctgct	240
35ttcttttgct ctgtctcttt tttttattta cttggggcct ttgagatagc tcagagagggc	300
gaaaggggttg gagataagag acggtgcgaa atagagggcg agtacgatga gcgtggataa	360
aatgcaggat gaaaagggtt agcggagtgg gagtgagggg tttgaagagg ggcttctgga	420
ggatccgaag gcaacgagta ggttgttgtt caagatcgat tgtcggtatg tttctctctc	480
cattcctgct ctccatgtct ttatcttgag ggcttttgtg gatgatgtac catcctgccg	540
40gttctcgccc tgctgttcct gtgctcgttc attgatcgta caaaccttgg gaatgcgaag	600

```

attcttgggtt tggagaatga tctccatctt acggaccacc agtacgctat tgggctttgc      660
gtctttttacg ctacgtatat tgcgaggtaa gcttcctgta tggcagatgc agtccagaag      720
actaaattttg tgcagcgaac tcccggtccaa tttgctgctg aaaaagggtat cgccaaagat      780
atgggttacct tttctgacag ccatctgggg cgtcctgacc atgtgcttgg gatttgtgac      840
5aaattttcgcg tcttttgctt ctgttcgcgc gctcctgggc gttgctgaag gaggcctatt      900
gcctggaatg gtaagatttt ggcgacgtaa taaaccgtct ttcgctaacg ccttgctagg      960
actatatctc tctcactttt atcgccgcca ggagctcgct ctacgcatag gcatcttcta     1020
tactgcagcc tctctatctg gtgcttttgg cggactcctc gctcgaggcc tcaatgccat     1080
tggcccagca agcggactcg aaggctggag atggatcctg atagttgagg gcttgataac     1140
10cgttggcgtc ggcgcatgct ctgctatctt ccttcccaat tccatcgaat cagccggttt     1200
ccttagcccc tccgaaaaag cccacgccc gcttcgactc ggtgaagcat ccgcctcgca     1260
cgaacgcttc gactgggccc aaatcaaacg cggcatcttc aacctccaag tctggctcac     1320
agccactgcc tacttctcta tcctctcagg cctctactcc ttcggcctct tcctccccac     1380
aatcatcaac aacggcttcg ccaaggacct caacaaagcc cagctctgga ccgtcattcc     1440
15ttacgcgctc gcttccgtct tcaccgtcct tgtagccatt ctctccgacc gcctcgctct     1500
acgtggccca gtcattgctgt gtacccttcc cgttgctatc atcggctacg gagtcatcag     1560
ccaatcgacg aaccgaaaag tacaatacgg aatgacattt ctcatggcta caggcatgta     1620
ttcctccgtc ccatgtattc tttcttgga cagcaataat tccgctggcc actacaagcg     1680
cgcgactaca tcggcgctgc agcttgcatg tgccaatgcg ggttggttcg tcgagcgtt     1740
20tacgtatcag aagagcgaga agccgaattt ccataagagt catagcatta tgctgggggt     1800
gttggtgctgc gcttgggttt tgtaagttct cttctccttg ttctcttttc tagtgtgtac     1860
aggtggattt ccatcgtttt gctggcggga tatgcagcta acgtgaatga tagggtcgca     1920
gcgaatgtgg cgtgggtgtg gaaaatcaac cgcgataagg cgagtggaaa gtatgcggaa     1980
ttcgaaggac gaggagatga tagggatccg gcgtttaaga tggatgatga agggattttg     2040
25gatctgggtt gggttattat tagcatgatg ata                                     2073

```

<210> 56

<211> 487

<212> PRT

30<213> Cochliobolus heterostrophus

<400> 56

```

Met Gln Asp Glu Lys Val Glu Arg Ser Gly Ser Glu Gly Phe Glu Glu
  1             5             10             15
35Gly Leu Leu Glu Asp Pro Lys Ala Thr Ser Arg Leu Leu Phe Lys Ile
             20             25             30
Asp Cys Arg Tyr Val Ser Leu Ser Ile Pro Ala Leu His Val Phe Ile
             35             40             45
Leu Arg Ala Phe Val Asp Asp Val Pro Ser Cys Arg Phe Ser Pro Cys
40    50             55             60

```

55

Cys Ser Cys Ala Arg Ser Leu Ile Val Gln Thr Leu Gly Met Arg Arg
 65 70 75 80
 Phe Leu Val Trp Arg Met Ile Ser Ile Leu Arg Thr Thr Ser Thr Leu
 85 90 95
 5Leu Gly Phe Ala Ser Phe Thr Leu Arg Ile Leu Arg Gly Lys Leu Pro
 100 105 110
 Val Trp Gln Met Gln Ser Arg Arg Leu Asn Leu Cys Ser Glu Leu Pro
 115 120 125
 Ser Asn Leu Leu Leu Lys Lys Val Ser Pro Lys Ile Trp Leu Pro Phe
 10 130 135 140
 Leu Thr Ala Ile Trp Gly Val Leu Thr Met Cys Leu Gly Phe Val Thr
 145 150 155 160
 Asn Phe Ala Ser Phe Ala Ser Val Arg Ala Leu Leu Gly Val Ala Glu
 165 170 175
 15Gly Gly Leu Leu Pro Gly Met Val Arg Phe Trp Arg Arg Asn Lys Pro
 180 185 190
 Ser Phe Ala Asn Ala Leu Leu Gly Leu Tyr Leu Ser His Phe Tyr Arg
 195 200 205
 Arg Gln Glu Leu Ala Leu Arg Ile Gly Ile Phe Tyr Thr Ala Ala Ser
 20 210 215 220
 Leu Ser Gly Ala Phe Gly Gly Leu Leu Ala Arg Gly Leu Asn Ala Ile
 225 230 235 240
 Gly Pro Ala Ser Gly Leu Glu Gly Trp Arg Trp Ile Leu Ile Val Glu
 245 250 255
 25Gly Leu Ile Thr Val Gly Val Gly Ala Cys Ser Ala Ile Phe Leu Pro
 260 265 270
 Asn Ser Ile Glu Ser Ala Gly Phe Leu Ser Pro Ser Glu Lys Ala His
 275 280 285
 Ala Arg Phe Arg Leu Gly Glu Ala Ser Ala Ser His Glu Arg Phe Asp
 30 290 295 300
 Trp Ala Glu Ile Lys Arg Gly Ile Phe Asn Leu Gln Val Trp Leu Thr
 305 310 315 320
 Ala Thr Ala Tyr Phe Ser Ile Leu Ser Gly Leu Tyr Ser Phe Gly Leu
 325 330 335
 35Phe Leu Pro Thr Ile Ile Asn Asn Gly Phe Ala Lys Asp Pro Asn Lys
 340 345 350
 Ala Gln Leu Trp Thr Val Ile Pro Tyr Ala Val Ala Ser Val Phe Thr
 355 360 365
 Val Leu Val Ala Ile Leu Ser Asp Arg Leu Ala Leu Arg Gly Pro Val
 40 370 375 380

56

Met Leu Cys Thr Leu Pro Val Ala Ile Ile Gly Tyr Gly Val Ile Ser
 385 390 395 400
 Gln Ser Thr Asn Pro Lys Val Gln Tyr Gly Met Thr Phe Leu Met Ala
 405 410 415
 5Thr Gly Met Tyr Ser Ser Val Pro Cys Ile Leu Ser Trp Asn Ser Asn
 420 425 430
 Asn Ser Ala Gly His Tyr Lys Arg Ala Thr Thr Ser Ala Leu Gln Leu
 435 440 445
 Ala Ile Ala Asn Ala Gly Trp Phe Val Ala Ser Phe Thr Tyr Gln Lys
 10 450 455 460
 Ser Glu Lys Pro Asn Phe His Lys Ser His Ser Ile Met Leu Gly Leu
 465 470 475 480
 Leu Cys Ala Ala Trp Val Leu
 485

15

<210> 57

<211> 1900

<212> DNA

<213> *Cochliobolus heterostrophus*

20

<400> 57

ctgccgacgg tagcttcgga gaatccaagt gtgagggcca tgctagcccg agaccggcat 60
 tgcgctaatt ggaccctggc ctgtaacgtg ggaaggacga acagcacagg tgcaggcttc 120
 tagggctgca tgcagtgcgc atcatctgca tgcacttgct gtgccaagtc gtgtactaca 180
 25caagtgcgag ttgctatctg taacgaggaa ccttgtatct aaaagtgtat acgtgaggta 240
 cgtgtgttcc agacctcaa atctaaagct actaaaacaa tagaaacagc ggagtctact 300
 ccgacaaggt caagtgaaag ggcgcggcat aaaagtcaat cgaatcaaag tacacggaca 360
 tacgagcaat ctacacacgg tcatggctat agcttacttt cgttctgctt caatcgatg 420
 acgccctatt catgtaagca cagtctacta tagcagacat aagcaagctg cttacctctt 480
 30ggacgcagct ggcaatgagc gtgccatcct tggatatacat tctctgggaa acgaggccgc 540
 gaccatcacc agcccaaggg gtctccatct cggatgaagat ccattcatct gcgcggaaac 600
 tgcgaggatt gtgaaagtag atggtgtggt ccagactaac catcatgcca atctcaggct 660
 ttgcgtctcc gctctttgcc aggtcttctg ctttcctcaa ttcacgatg cgctgcttgt 720
 ctgattcggt gacaaagctt tggcgctgta actcggcatc atccatctcg agcagcttct 780
 35taagtacgtc ctcgtcgatg ctcgacctgg ccctgctctt gcgctggttc gagtagcgca 840
 gaagcttgtg cgcacgcgcg acggtgccga tgaagtagct atcggacatg tatgcgatgg 900
 cggagagatg ggcttcgtga ccgccagcgg gggagatttt accgcgagcc tttatccatt 960
 gtcggcattt cttggtgtgg ggcttgtcgg agtcgtctgc tagatatgag ctagggcagg 1020
 cttaaggatg gtatgcgaag tcactcaccg ttttcaatgg gcaacagctg ggtctggaag 1080
 40ggactctggc catcgctggg cgtcttcaag tcgtcgctac cttccttggg cgccgggacg 1140

57

```

tctggcatcg ggtagatgtg ctcgaccttt tgagcgctc cactgttctg gcgaacaaaa 1200
ctcatgggtcg tagtgaagat gacgttgccc ctttgccggg cctgcaccgt cctggttgcg 1260
aacgactttc ccgagcgcac cttttctaca tggatatatga cggggatctc ggagttgcct 1320
gcaaggatga agtagcagtg catcgaatgc acagtgaagt cgggggtcaac cgtcttcttg 1380
5gcgccgctga gtgtctgggc aatggcagca ccgcaaaga tgccgcgcgc accgggggga 1440
tgccataggg gacgagtgtt tgtgaagatg ttgggatcaa tgctggccag ctgcgtcagt 1500
tcaaggacgt tctcaatggc cgactgggag tggtcggcgg gcggggggcg gatgagggtg 1560
gccatgggtg tggctgatag ttttcctgtt ggtggatcgt tctgtgttct gcgaaaagga 1620
ggccagtgtg gcaagaccag atgcaagcag cagcagcgag cggctgtgtg agactttggg 1680
10cgtcgtcatt tccggggcac gtcaaagcag cgcagacgcg catgagccga ggcacaatga 1740
tcatcggccca tgtgggagct tgtcgcgccg aacacgtgac tggccgctga ctgatggggg 1800
ctgactaagc caggcggcgc caagccgagg agcaggctgg ctctggggta aaaacgtcat 1860
actgggcttg ccgggccctg cgcagatgcg tacctggctt 1900

```

15<210> 58

<211> 368

<212> PRT

<213> Cochliobolus heterostrophus

20<400> 58

```

Met Ala Thr Leu Ile Arg Pro Pro Pro Ala Asp His Ser Gln Ser Ala
  1             5             10             15
Ile Glu Asn Val Leu Glu Leu Thr Gln Leu Ala Asp Ile Asp Pro Asn
      20             25             30
25Ile Phe Thr Asn Thr Arg Pro Leu Trp His Pro Pro Gly Ala Arg Gly
      35             40             45
Ile Phe Gly Gly Ala Ala Ile Ala Gln Thr Leu Ser Ala Ala Gln Lys
      50             55             60
Thr Val Asp Pro Asp Phe Thr Val His Ser Met His Cys Tyr Phe Ile
3065             70             75             80
Leu Ala Gly Asn Ser Glu Ile Pro Val Ile Tyr His Val Glu Arg Val
      85             90             95
Arg Ser Gly Lys Ser Phe Ala Thr Arg Thr Val Gln Ala Arg Gln Arg
      100            105            110
35Gly Asn Val Ile Phe Thr Thr Thr Met Ser Phe Val Arg Gln Asn Ser
      115            120            125
Gly Gly Ala Gln Lys Val Glu His Ile Tyr Pro Met Pro Asp Val Pro
      130            135            140
Ala Pro Lys Glu Gly Ser Asp Asp Leu Lys Thr Pro Asn Asp Gly Gln
40145            150            155            160

```

58

Ser Pro Phe Gln Thr Gln Leu Leu Pro Ile Glu Asn Ala Asp Asp Ser
 165 170 175
 Asp Lys Pro His Thr Lys Lys Cys Arg Gln Trp Ile Lys Ala Arg Gly
 180 185 190
 5Lys Ile Ser Pro Ala Gly Gly His Glu Ala His Leu Ser Ala Ile Ala
 195 200 205
 Tyr Met Ser Asp Ser Tyr Phe Ile Gly Thr Val Ala Arg Ala His Lys
 210 215 220
 Leu Leu Arg Tyr Ser Asn Gln Arg Lys Ser Arg Ala Arg Ser Ser Ile
 10225 230 235 240
 Asp Glu Asp Val Leu Lys Lys Leu Leu Glu Met Asp Asp Ala Glu Leu
 245 250 255
 Gln Arg Gln Ser Phe Val Asn Glu Ser Asp Lys Gln Arg Ile Arg Glu
 260 265 270
 15Leu Arg Lys Ala Glu Asp Leu Ala Lys Ser Gly Asp Ala Lys Pro Glu
 275 280 285
 Ile Gly Met Met Val Ser Leu Asp His Thr Ile Tyr Phe His Asn Pro
 290 295 300
 Arg Ser Phe Arg Ala Asp Glu Trp Ile Phe Thr Glu Met Glu Thr Pro
 20305 310 315 320
 Trp Ala Gly Asp Gly Arg Gly Leu Val Ser Gln Arg Met Tyr Thr Lys
 325 330 335
 Asp Gly Thr Leu Ile Ala Ser Cys Val Gln Glu Val Ser Ser Leu Leu
 340 345 350
 25Met Ser Ala Ile Val Asp Cys Ala Tyr Met Asn Arg Ala Ser Tyr Asp
 355 360 365

<210> 59

<211> 42115

30<212> DNA

<213> Cochliobolus heterostrophus

<220>

<221> misc_feature

35<222> (1)...(42115)

<223> n = any any nucleotide

<400> 59

gatcttcttc acaatagtcg tctttccgc attgtccaga cccctagtgc tgtcagtcct 60
 40tgccacaacc attcttctgc aaacacgcac agcatcagga tgccatctc cttgtccttt 120

	aagcgagctt	ttcgtaatat	cgaaagcatc	ttgcgctgct	caaaatctga	gaaaatggtc	180
	ctactaggca	gaagcaagac	agtataatg	ggcttcccag	agccaccttg	gagctaagcc	240
	gtttgcgga	gcctgcattc	aacgccaact	cgctacgctt	ctttggagac	aacgtctttc	300
	cttgtcagat	gatcgacact	gcgttgatga	ctcgaaccag	ttgggagggtg	tagttcctcc	360
	5tttattttat	tgagacatca	tggcgacagc	tgtattgctt	gcccgtatgc	tctttcctac	420
	taacagcacg	attgcttact	atgtgaagac	tcgttggaca	tgagcgctt	accaacaaat	480
	actcccactc	tatagaaaga	agtcgaggta	aagtgaagtc	aagtgaagtg	aacaagcatt	540
	cactgtatgc	tttaggcagc	tccgaccaag	tgtatagaag	gctgcatcat	ctgccattcc	600
	acctctccac	cttctgcttt	tcgccgatcc	ctcttgcttt	acttagaacg	ctcctgatat	660
10	ccgttccctc	tacaaaatcg	aagaaggctt	gtatcaggaa	cacagccatg	gaggactcgt	720
	gtctttgcac	ttcactctca	cagccaccgg	agcttgaact	gacaagattg	ccatctttcg	780
	ataccacac	caatatgtcg	aagcaacaga	aaaacagaga	tatgacttgg	cagcaacagg	840
	cagacttcgt	gcaggccaca	gtaagtgaag	taagccttat	gaccgcatca	tgtctcacta	900
	tcctttcgta	atcatacacc	aaccgtgggtg	aaagccacag	tagacttttc	tagcaccctc	960
15	accattcccc	gctctcaaca	acacctcatc	atagttttct	attactcact	acctacccca	1020
	tcacccccca	ttttaacata	tcctgcgctt	gtatgctaac	aacccaaacg	acgaaacaga	1080
	ggccactcaa	acatcacttc	cgtcgacctg	ccaagaaacc	accacatcac	aaaattaata	1140
	agaaagcgga	tgcagtagca	aaagaaacgc	aaaacagcgc	gccgttacag	caaaccgagt	1200
	caccatccca	gcctgactcc	agacgcaaac	atcgtgcgca	gcagccaact	accttaccac	1260
20	ctcctaacca	cgtctgtcca	gctgcacttg	ccgtgacatc	gttgcccagt	caagacaaga	1320
	ccaaaaacgc	atccgaccct	ccagtccctt	tgcagaaaaa	acatacgcg	aacaagaaga	1380
	gaggtaaaaa	agaggtaaac	atgactacgc	tacgaaaaga	agcctatgtt	cctccacacc	1440
	tgcgcagctg	tccccctgcc	aacaaggcct	ctattccgcc	acacctgcgc	agccgccctt	1500
	ctgccataaa	agcgacgata	gatccagggt	ataaaaaatg	tgccaccaat	ggctcacctt	1560
25	cttcttcgaa	aaattccaag	tccacagttg	ctaccaagcc	cgagtcagtt	cagtaagtcc	1620
	ttcttgacat	tacagggttc	gactgctctt	ttttgtatct	tcaccattct	gacatgatcc	1680
	aagaaacca	aacaacatgc	gcgaggcaac	accacnttca	ccagcaacca	caccacatga	1740
	gcctgtagag	catgaacata	aagacatggc	tacccccgaa	aacgtctggg	gcggctggaa	1800
	tgaaactgaa	atcaacaatc	tgcattgctc	tcagaagacn	tgctaaaccg	cgttggaaatc	1860
30	gtggcactca	gccgtacaag	aggaagcctt	ggccaaaaca	aagggacatg	aaatatattc	1920
	ctggcaaaa	cgagagcgat	ggtggtgggtg	tcaactgctg	gtctgacagc	aacggagacc	1980
	ctgactacga	tgtcaggaaa	ctgctagact	ggaacggcga	ttggctacct	gctccggaat	2040
	catgggtccg	tcgaagagga	catgaagacc	gtcaccttgg	tgcacatgta	gaacaatgga	2100
	tgaatggaca	ctcacaagag	tgcaccagat	ccgtataacta	cccactcagt	actttcagtc	2160
35	ccgaagatgg	accttgcaaa	gagctggcac	ctcgttactg	gcttgaggcg	aagggtgagg	2220
	gcagtaactt	gagagaatct	tggagacaaa	tctctaactt	ggacccaaa	ccgctggatg	2280
	atacggacat	tactatccat	ccaccttgggt	gggaattgta	cgaggatgtg	gtctattctg	2340
	aggtgattca	cgaggaagggt	caggggtgaac	agcatttcaa	gcataggagc	tggtacctga	2400
	acagcctacc	agcgccggag	gcaagaatcg	accctaccga	tgcagagcat	cctaccactc	2460
40	atctgatgct	ggcttcggct	gcagaaaagc	ttcaagatct	acaacaacgt	aggggaagcta	2520

	aggaacgctcg	cttggtggcc	aaacggaatc	gccagtcgc	gaattcgatg	tttccaatgc	2580
	aagccatgga	agatcgtcgc	ctacgcccta	agaccaacat	gtacattcgt	cctgttcagc	2640
	cagcagatgt	tgttggcatt	ggagtaagtc	tgaacttaca	tagttcctga	ttgacttgga	2700
	aaacccatag	acaaggatgc	aaagttttca	aactaacaat	attgacaggc	gatttacaac	2760
5	tactacggtt	agcataccat	ttacgcaacc	gagtttgatg	ggcgcaactga	agatcaaatc	2820
	cgccagcgaa	tcaacactgt	caccagtgca	ggccttccat	acttggtcgc	agtctcaaag	2880
	agcaacgagt	ccaggaccaa	tcccggttat	gttaccgaaa	agattgtagg	cttcacagc	2940
	ttggatgatt	actgcagcca	ggcatcctcg	ttccgctaca	cttttgagat	ggagttgttc	3000
	gtccacccag	gctatacgag	caaaggtatt	ggcaagtgtc	tctggtgatcg	tctcctagag	3060
10	atggcggaca	caagctaccg	cgctcgcggc	gggtatcagt	acgtcaacaa	cttcgagtac	3120
	ctcaagaccg	ggccatcaag	ggttatcaaa	acgattctac	tcaacgtcca	ccacgagaat	3180
	ggagagcatg	cagagaccgg	atggcagggc	cagtttctcc	acgcatgcaa	gtttcatcgt	3240
	gtcggtcggc	tccccaaagt	gggatataag	aacaacactg	ttatagatgt	tgccatctat	3300
	gcacaccaca	ccaacgaaga	gattgatgca	ggtaccgcgc	ctactgtcgc	aggataaccc	3360
15	agctcaacat	gctgcttgac	gaaaggtgag	ttattcaggt	aggtgttaga	atgagactga	3420
	ctaaggattc	agatcatgta	acggttggtt	tttactgtc	ccggtatctc	atgcggcaaa	3480
	agcagttgca	caaacggaat	tgtgtcattc	tactcctatc	atctgctgta	ccggcttggc	3540
	aacagtggat	tacgaaattg	attctttgct	tttgcattga	ttaaccatac	gcattgaagga	3600
	ttccaagggc	aatgggtgtc	agagatcctt	gttcttcgat	gtcctcttac	ttttgaagga	3660
20	attacgtatg	acggatgatg	agaaccgtca	ctggtatcag	attgacccaa	acttgaactt	3720
	ttcggggctg	caaattgcaca	cttgctccga	cactgtacag	tagatttccc	ctattttcaa	3780
	accgaatata	tcgatactca	aatgaacatt	gtgatgattc	tgcatgacga	ctgaaatggt	3840
	gtcgatacct	tgtcgcgtcc	catacccac	taatcctgta	acgcgtcgac	gctcccacgc	3900
	tcatgaagcc	accagtgcgc	tcacagtcgc	ccccaaagct	tgatctgcgc	agtaacacat	3960
25	aaccacgcgc	cgggccagtt	ggattcgagc	tgaacagaca	tcaagtcac	aaatctacat	4020
	gtgagtgtgc	cttattaaat	attcctatct	tcccaactca	taccaaccac	aacccggaca	4080
	tttacagatc	tgtatctggt	aatccaaata	ccagacaacc	atcagacatc	gacagagtct	4140
	ttgaaagaaa	tcgtgtaaca	caacacttcg	tgccaaatcg	aaagtaacaa	aagagggacc	4200
	tcaaaaaaaaa	acaccaact	cagtcaacaa	gaaacacgaa	aatggcgcaa	gagaagaagg	4260
30	aagaacaacc	ccagcaagac	cacatcccca	cctcgcgcga	gaacgaagag	gaggaacaaa	4320
	gcaaaggctc	cggcggcctc	ttgagcgcaa	tcggagatcc	agtcggtacg	tctccttacc	4380
	cccccttccc	cctccatctc	tcaaccacac	acctaaccac	tctcccaagg	caacgtcctc	4440
	aacaccgccc	tccgccccgt	cggcgcgcgc	ctcgagaaat	tcgtcacagg	cccgtggggc	4500
	gagggctctcg	gcggcaccac	acgcggcgcg	ctgggcccgt	tgatgggcca	cgaggacgag	4560
35	cgctctgagc	tgctgggcgg	caagaacgta	gatagctaca	gcaagcccga	gaagattgcg	4620
	ggtaagggaac	agacggggaga	taatccgttg	ggcttgatc	agacgggtcg	atggggattt	4680
	gaggatgagg	gtaagaaata	gaagagtttt	tgtttgattt	taagaaagtt	aaaagtgagg	4740
	aggccggggg	aggggggtata	tataaatctt	ttttgtatgg	agggaggaaa	ggaggaaatc	4800
	aaaacatttc	actcatgcca	ctatctccca	acacaccttc	ttcaaagtac	tcgtgttcct	4860
40	catgtcctcc	atgttcttga	tgtgcatgtc	gccgcgatcg	aaacgtatca	tctagctcct	4920

gcactgtgcc	tgtactatcc	ccaaaccccc	cttcctccat	gctatctctt	gtacccggta	4980
cgggtgtccg	gccagtcagt	atatcataca	aactatcatc	tgatcccgc	tcctcgccct	5040
cctcttcgcc	ctcctcttca	ccctcttctt	cgtcttcatg	atcaaaccgc	agaatctggt	5100
ctttgccctc	cggccaaacg	acaacggaag	cttgagtatg	cgcgagcgcg	aaacaagggt	5160
5ttttactact	agctgcaggg	cccagagacc	aggatacggg	ccagcgtgcg	ggagtcgcga	5220
ctgcggggtt	agcaatgtga	ggggataacg	agaggatgga	cgggtggatgt	tgcgtagtgg	5280
ccgatgaatt	cgccgatgaa	gatgacgtcg	agtgggagcg	ctgtgaagcc	gtgtacaagt	5340
acacgactgg	ttcgtcatgc	gcagtttgga	taacgagacg	attgggggtcg	caggggtgcc	5400
agaggagtgc	tttcacagga	gcgtacatta	tgaggatcga	gcggggacgg	agacttcgca	5460
10ggtcccaa	ccagactgtg	caggggtgtg	tgtcgtctct	gcttgcgcac	attgtgcctt	5520
cggagttgaa	gctgagcatg	ccgatgcctt	gttttaggag	cgcattttcg	ttcttttcta	5580
gggcggtttt	gggaggtgtg	gcgggttgtg	gtgtgagtgt	gaaactacgg	gcgcccaggt	5640
tgtcgacttg	ctctgtgtac	actgggtgcg	tgggtacgtc	gatgacgggt	gtgtggtcga	5700
ggaacaggat	gggtgcgaat	gtcgtgttag	aaaggatgcg	aacacgacgg	tcccagccgc	5760
15caactgcgag	acgttcatgt	ccagggaccc	attctagact	cttgatgcct	aggccttcta	5820
catcccattc	gctgacgtcc	tcggatgctt	cgcgggttat	ggtgcggtac	aaatgcccat	5880
ccgccgtata	tatcaaagct	tgtaaccgcg	agacgcagcg	tcccagatgg	ccagccagcg	5940
cccgtcacga	ctccatctca	gaccagcggc	gtctgtagta	gggagttcga	ctcgattcag	6000
aaccttgtag	gtctgcggtg	caagaagcaa	caagatatcg	gtccctgatg	cangacacaa	6060
20taatgccaga	acacgccctt	gtcccccttc	attcctcaat	ccagtatcgt	cagcaggtcg	6120
gtaaccccac	cccttgccat	ctttaccagg	aaacttcgga	tcgcgtatct	ccactacccg	6180
accggtcttc	aagcaccata	tcttaacaca	ggcggtaaag	tcgggtccaa	caagcacctc	6240
gtcctctgtt	cctccaaact	cgacgtgaac	attcttcccc	atgccaccag	agccattgct	6300
aatcacggca	ttccatttct	catcgcgagg	atcgtaaacg	cgcgcggtgt	cgtcgtcgga	6360
25tatgaggacg	cgattcgagc	agggacgtgg	tgtccgtgat	gatcgacggg	gaggtgtagt	6420
ggtgggcgaa	gatgtgcgtg	ttgatgaggt	caagggcgga	atgaccaggg	gtgaccaggt	6480
aatcttcgac	gagcgcaa	catgggtgga	tgggagggcg	atagtacgga	ccacctcgaa	6540
agtattgaga	catcggtatt	gcaaacgtgc	accgttgaca	caggcagtat	gtgtcgcggt	6600
cggcgaggga	acagacaatg	tcgtggctgg	tgcgaatcag	taactgctgt	gtaggatagt	6660
30ctgtttatta	gacgcacatt	tgatttgctg	ggatatctcc	atgttctcca	ttgcgctgga	6720
cactgacggg	cgtctcaagt	ggctttgtca	tggggatggt	tgtgtccggg	aaacaaaagg	6780
gagcgacggg	cgttcagcca	atgagcgttc	gaattggccg	caactagcgt	gaacgctgtc	6840
cgcattggcct	ccgggcttgc	tcgctcatat	gtacaggcct	cgtgggttaa	atagctcatc	6900
tggttcaaca	gacgcattca	gagtcattgt	aatccgagcc	aaggacactg	tgtttcgtag	6960
35gtccaaagac	ttgatttcta	ccacacccan	acgcaccaac	agtggggggg	gtttgtatgc	7020
acatcaaaat	acaacaaaaa	aggtatatgg	acaggcatga	aacgtactga	atacagtctt	7080
cgagagaaag	catatccagt	atgaaagcat	ggccgtgcaa	aaaaaaactt	gtatatgaag	7140
aaatggtgaa	gaaaatgccc	aaacgcttgc	ttgccaaatc	actaaattcg	aaacatacat	7200
caaccttcat	cttcacgtga	aaaacctcaa	gaaagctcat	gtgctgtaca	caggagcctt	7260
40ggtcaagact	gttttgccgc	gaatctgcac	tgcgagtccg	acagccaagc	gtacgcgggg	7320

	tggtgcagac	tcgtctttcg	taccagtcga	gctgcctggg	gcgtttttgtg	tcagactctg	7380
	ctggctacct	ccttgaccaa	gagaagattc	ccgcttgccct	agggaagcag	gagtgcactcc	7440
	tgccttcttg	gacgtggcga	aaccaaacat	gctgcgtttc	ttgtcctctt	ttgaagtctg	7500
	accaacttgc	ggttgaattt	gatcttgctc	tgcgacggag	acgccatcat	ctggcttcga	7560
5	tgccatgctt	ctgatggaaa	ttctgtcagc	atcatccagc	tcgactatgg	gagctgttaa	7620
	ggcgggtatg	ccggattcca	gcttcacaa	gtcgtccaca	atgccatgac	ccgggtggaaa	7680
	gtactcaatc	tgcacttctc	ggctttctaa	tggcaaagt	tgggcaatac	tgatagcaag	7740
	ctcaatgatc	attggaacac	ctccattcag	atccggaggc	gcggggtcat	tcagatgaat	7800
	ctggagagat	gcgatcagtt	tttcgttgag	cgtttcggtc	agtttctctc	gatacgtagt	7860
10	cgcttgtggg	gctttcagac	tatccgcaag	gccttctagt	gtagcaagcc	gccagctgac	7920
	aatcttcgca	gcaagtgcact	cttcactctc	tggactatgg	caagggggcg	agaacttccg	7980
	gatattcgac	acaatcgact	gtagtttgt	cgaaaacgct	ggctcaagat	ccggatgaaa	8040
	gtatttatca	aatatgttct	ccactagcca	ttgagagatg	aatgcacgac	cgactgcagt	8100
	catttcctgc	ttgcccgctc	ccacagctgt	cttggtgcact	actgggtgca	gccactgtgg	8160
15	tattgacttc	cagtccttgc	gaatggaaaa	ggacagttgt	gctattagtc	cgtctaagcg	8220
	attgaagcgg	gttgaatatt	cactgtcatc	ccacgatgtg	cgcgaaacag	ataaacgctg	8280
	gtttgcaagg	gtattctgca	gttgggtggac	ctgggtttgt	tgttcgaaaa	agtatctttt	8340
	gaccttttga	tattttttcac	ctgcagcctg	tcaaactcaa	tgatctttct	cttgagatat	8400
	tgtacttacg	taaaacatca	tgatccttta	tcatcttttc	aatttcctcc	tccgtcatth	8460
20	cggacgcaga	tgtgcgcggg	ggaggtgtcg	cccgcctcat	ttctgagcct	tggtgacctt	8520
	gcttgtcagc	tcgatatggc	tgtccttggg	aggctgtggg	tacgacgttg	gcgtcgaacg	8580
	ttgcgaggcc	cggagacaca	ggcggctggc	cgaaggtctg	gccttgagcg	gacgggggct	8640
	ggagctgaga	tgactcttgc	gacggggctc	ctgggttgga	ttggctactc	tcgttcactt	8700
	tcctcacagt	gcttctgctg	ctacttttag	cacttggtgt	cataccctcc	tgtccttgc	8760
25	gatgcgatgc	gttggtcagt	gaatgtagct	ccgaagatag	agggtgcttt	ccctgctgcg	8820
	attcggtgtg	ttggtagggc	gtgggctgaa	gaggcgacgg	cggggcaagc	tgatgggacg	8880
	agggcggtcg	gcgttgctgg	gactgctgtg	gctcgattga	ctgttggtaca	gtctgaggat	8940
	agtgttgctg	tatggcctgt	gggtgggtgca	gttcgggtctg	cagttgcgcg	atcggcagag	9000
	cagtttgagg	ctggcgctgg	tctgggttag	gctgaagttg	agatacaaa	tgaggctgct	9060
30	gatgatgcag	gctttgcggg	gaattctcta	gcgatgattg	attatgcgct	ggcggaggct	9120
	gttgctgctg	ctgctgctgg	tactggtaat	aagttctgtt	gaatgtcttg	tactggagcc	9180
	ccttgcccat	gcggctgagt	ggacaaagag	ccttggcgct	tttgtggaat	gtacgtctga	9240
	aagttgcctg	accccggtga	actctgggta	gtgggctggg	agctgtgctg	tggctgcttg	9300
	cgggtgaactt	gctgttggtg	cggttggttg	tgttggtgtt	gttggtgctg	ctgctgtgga	9360
35	ccttgacatt	gttggtgttg	ctctggggcg	tactgatcta	ccccgccttg	ggtcgaatag	9420
	ctgccctctg	tgctcaccct	gccaaagtgg	ggccgaacgt	aaggttcctg	gtgcaactgc	9480
	gcagcgaacg	gctgaggggt	ttgcacgttt	tgttgatatt	tgggtgtctt	gggaggcact	9540
	tgaggggact	cgctgccttc	ctgctggagg	aatggatcaa	ggttgtcctc	gtcttgctct	9600
	cgcgatgtgg	gcaagtggga	agaggacccc	tgggtgtgat	gccaatgtgg	cggcggtctc	9660
40	ctgcctgtgc	gtgatactcg	tgcgcccag	ggtcgctttt	ccgtaccgac	tgtcgcctgc	9720

	ttaggccggtt	cttctgagct	ttgggtctcct	cgtccttgtg	tcctctgacg	cccgcggcaa	9780
	tgcgactgcg	cagcgactgc	ttcttggtttt	cggggctgct	gtatgcagga	ggaggagagt	9840
	gaggctggta	ggggacgtgg	tttggcgatt	cgtggcctgc	ctgagatgta	ggcggcgggc	9900
	taggctgttg	ctggcagttc	tgcgatggct	ggcctaagga	gaggtgtgca	ctcagtctgt	9960
5	gagttgcatt	tcgcagagag	tggtattcgc	tgtattggcc	ctgggtgctt	ccagaagagg	10020
	ccggtgggtt	tagagaagac	tgctggctat	ggacggggcg	ttgatgggcg	tcggaggcct	10080
	ggatcgcgtg	ggctgctgcg	gcttggctgc	tggcttcctt	tgcagatggc	acgggcgaag	10140
	gttggctcct	tactacggtg	ttatcgacag	agtgggtgca	cgggttcgac	tttgtccatg	10200
	gaaagttagg	cattgcaatg	atgacagctc	ccagttccgc	cgcaagtat	gttgatgatg	10260
10	gcggggcgga	gggtgatgcg	cccagcta	aataaaccaa	gcctgagcaa	cggttaacct	10320
	aggcgatgat	tgcacccgaa	cgagacagca	ttgcgcgcgg	ttggacagct	gtcttgaaaa	10380
	ggatgggatg	ggctacatag	tagatgcgcg	tcttcgcgca	tcagccccag	ccctgcccgc	10440
	acgtgcaggg	ctgatgagca	aatgaaacca	gatgcaggcg	cgagtcccaa	ttccgggtgca	10500
	gctaactgca	caaggagatg	ctgatatggc	ggcgggtggca	gtgctggcgt	cccgcgatgc	10560
15	tgcttttgta	gtctaccaca	ggcatgttac	tgtgctatgc	cgtctagtcg	ctgtcaaagt	10620
	gtactagttg	aacgctgtat	tatcatttgt	catggcggtg	gcaacgcacg	acaagcacgc	10680
	caagtggggg	aacagttcat	gtacgtagat	aagtatgacc	cactggagat	ttctgcctcg	10740
	aggagacacc	cagccttgcg	ctccattcct	cagcctgctg	tctagttacc	gaaggcgag	10800
	agcatacacg	tgctacgtct	ccacgagaga	ggtaagcagt	cctcctcagc	tttgcccata	10860
20	tccatcacca	ccaaatcttg	atcaaacc	cagtgtctta	cagtcaaaaa	aagtatcaca	10920
	ttacctaaac	acgactttcc	aaaattccat	ctcacgtatc	ttggcctatg	ccggcccttt	10980
	gacgcgcagg	atcagctcga	cgcccgactg	ctcctaacgc	tgcgcaggca	acaaaccggc	11040
	ccttagtata	acaggcgtcc	acgatcaggc	cgccagcctg	agtcaactaa	aaagcaccgc	11100
	tgctgtattc	tgcacaaggt	aagcaaaaag	cgtctcgaaa	gcttgacgtc	gcaaggcggg	11160
25	gggcttattt	gaccatacnt	tactcttacc	acttgccgca	acatcatgtg	tcgcgcgctg	11220
	tgtatgttgc	aagtatggta	agcacagcgc	caatcggatc	tttacctcaa	ttttacgccc	11280
	tccatgacat	tttgctggct	caggtttccc	cntgctgctc	cacagttagg	cccagtctac	11340
	tctattttcg	gccaccggac	gtgtgctagc	tttactccat	ttgctaggca	tctgcgatat	11400
	tactcagtcg	tcttcttacc	tagccttcag	tgctggattt	gatatcagat	acttgctcac	11460
30	cctaattcgt	cactctgact	actggattgt	atctgggctt	cacgattgca	ttgggtgattt	11520
	tcaatgacta	aaaaaagggtg	cttgccgggtg	tttcaaaaagg	aatgcatgtg	tatgtcagag	11580
	cggttatgca	cgcttccttc	agattgttgc	ctccaaaaaa	ccaacatatg	cagcattgcg	11640
	gcgctgggca	aggaacgccg	caacaatgtg	acatcgcgca	gcgcttccta	atcagcctta	11700
	cactctcact	aaatcgtcaa	gatcaagaac	gaaaggggtcc	aatgccaaac	gtggatatcat	11760
35	ttcctgtgtg	gaagaagcat	gacttcgtaa	atcaagaggt	tgatgtacat	attactgaac	11820
	ttcttcaaca	taacccaacg	cctcgccaac	cttgactttt	tgcccttggt	cgatattcca	11880
	gtgaaaccca	ccttttctct	cgccacgtgt	accactaaag	ccctcgtcca	aactaggctcg	11940
	aatgcccttc	ggcgcttcaa	agactaagac	aattgtactg	cccaactgaa	aaccacccat	12000
	ttcctcgccg	cgcttgagtg	cgtaccctcc	caagacacgg	cttgcgctcg	tgtaggaggc	12060
40	ctcagcgaat	ccagaatacg	gctcaccacg	ggcagcggct	tcttccgcag	cacgggtccgc	12120

cgcagtgtcg	gttgtaagc	tgtttgtgcg	aagttcgcga	tcaaagttga	tcttaatgga	12180
accaacgttg	gttgcgccga	ccggagtgtg	ggaaaagaaa	ccccagcgcc	atcttcctag	12240
gagaaccaca	cgctcgttca	gggtaaagag	accaggcata	gtgcgttgta	ggtagggcga	12300
tacactataa	agctcgccag	caaagtgacg	acgcgactca	acaacccatg	atacagggtga	12360
5gtggaacctg	tggtagtcgc	ctggcgcaag	atatacaacg	cagtagtaga	gaaccgtagg	12420
tgtctttaat	gaggcgggtg	cccacccatg	gcgctgtgat	tcactcaagg	caaggctcggc	12480
acgtacttcg	gcttctgacg	atggctttga	cggaactgat	tgatccgtcg	gcatttcagc	12540
aggcttcccg	tcttttggcc	atggctccga	gaagagggtt	ggtagagtat	atgagatacc	12600
gttcacgttt	gcaaattcct	catccgcgcg	cacagtgtcc	tcttcgtctt	gtgggtgtctt	12660
10ctcgtgctca	ctagcgcgaa	tttgggaatt	tgctacattt	tgctctgggtg	tactggncct	12720
tgtagatcct	agcagagcgt	ccaaactata	tgttacacct	ttgacttgct	caacttcgcc	12780
gtgctcgatg	gtgccaaatt	gaatgatctt	gccgtctgcg	ggagagagta	ctgcgttggg	12840
gttgggatct	agaggacgta	caccgggttt	gagggtgcgg	tagaaaaagg	cggcgagggt	12900
ggggtatata	tgtagatctg	gttccgagac	tccggagaga	ctaggaaaag	ttagatacac	12960
15ggacgatgat	gattatgggg	cgacttgctt	gacaccaa	atccaagaat	acagcttgaa	13020
tccaggcaca	cgaaggtagt	agggtatgtc	gatctcattg	aagcgacccc	acagtcgcga	13080
caacgccttg	agaggaaggg	tagacatgac	ctgaacggtc	catggtccgc	tcggtcttat	13140
tctttcgcgc	ttcttcggac	gaccttgctg	atccactaca	ttgccatccg	catcccttct	13200
ctcggcttct	gtatgctttt	ctctgcgttg	tatgcgatat	agctgaaatg	caccaggtaa	13260
20gccaataaccg	agtgcatttg	gtattggctc	ccatttgact	ttggtattct	tgagtgcgga	13320
attgagccgc	gacctaaaag	actccttctc	gtgaaaattt	tgttcgctat	atgacgctcg	13380
agtcgacgtg	aagggtgcgag	atggggcaga	gaggcgcgca	tgtgggtgtg	tggttatatt	13440
gcatcgcggt	cggatagaag	ctcgaatgag	ctggcgcgaa	caggagatcg	ccatgttatg	13500
atgcatgcat	ttgcgatttc	taccgttggt	ttctcggcca	catcaagatg	aggagattcg	13560
25caatggcaaa	gcgggaagat	gacggcgaag	gcgttgggtg	gtgtgggaaa	atccacatga	13620
gcggctaacc	catgacgtac	ccactgggga	gggattagcg	ccattgcggc	cgtgggctcg	13680
gagggtgtttc	cacgagcatc	caacacttac	tgtcgcacat	catgtgcatg	tcgtcaagac	13740
accttgaata	tacaagggta	ggcggcggaa	agtcgcacat	gtaatatcca	gggcttcgga	13800
aggctcagcct	gataaactcc	tcttcactgc	acatggcaat	cgacgttgta	gttgacggcg	13860
30aatccggagc	tttactctaa	aacgcaacgt	atacaaacac	gtcgccact	cccaaaccga	13920
gaaccttcta	tcgctctcac	catagtcgct	gctgcacatg	gaaggcatgg	cggacgcaga	13980
gcagacaatc	aacctcaagg	tcctttcgcc	ttcagcggaa	ctagagggcg	gcatcaccct	14040
cgcgggccta	cccgttcta	tcacgggtcaa	agagctccgc	acccgcatac	acgatgctgt	14100
gccctccaag	cctgcccccg	agcgcacatg	cctcatatac	agaggccgag	tggtagcgaa	14160
35tgatgcagac	actctgacta	ccgtgttttg	cgctgacaat	gtatgttgct	actatggcca	14220
aatgggcgct	tgctaaccag	aaccatagat	acgtgagaac	aagaaccaaa	gccttcacct	14280
cgtcatacga	gagctgcctc	caactgcac	ttcgccgtgc	ccgcaatcgt	cttctgtccc	14340
accaaacctc	ttccgctctg	ctgggtccaga	tggcccgacc	gcgagccctc	tgagacgaa	14400
tccatttcgg	gctataccac	agacacgacc	ggcttcacaa	cctcaaatac	ccagtcgca	14460
40ccttccgcct	catcgccttc	cgggacaagt	gaacccatt	cccataccat	taccgcaca	14520

actccatcaa	acgtttgctc	aagcaatggc	acaccaagga	caacaggggtg	atgaacagcc	14580
ctcagatcga	actagcgagc	agccagatca	aggtacaccg	gcagcggggg	ataggacgca	14640
tacaccaatc	ccttcaggac	cgtcgaaccc	tcctggaaat	ggcgaccagg	cgatcaggcg	14700
agaagggtgt	gcgccaatg	gagcacgatg	gacagttacg	gccttcaatc	cacttaacat	14760
5agctgcgcga	ctcccgccgc	ctgtcgtcac	attccctgtc	ccgcatgcac	taacttttcgg	14820
tcgtccgcgc	ctttctagcg	acaaccagcg	gttattgcct	cgtgtgcaca	ggatcttctt	14880
ggagacaaaa	cgggagattg	ataacattcg	agcattgttg	caactgcctg	gtgcatctga	14940
tgcacagagt	ggagggtccc	tcacctcaga	tatacctgcc	tcgttgaata	tccttgtatg	15000
gcgaatcgag	cgactacgtc	agcacctgaa	cacagtcaat	caaaatctgg	atgtcgttga	15060
10ccgggctctg	gcgttgcttc	ctacagagcc	tgaagtgcgc	gcgctcaggc	gctcagctac	15120
cgagttgagg	gttgatgctg	cggaattgag	tattgtgctc	gatcgtcaac	agggcgaaac	15180
ggccagggtc	acttcggata	cagcaccagg	ggtgcccacc	atagctgcgg	catcatcaac	15240
tacatcccag	acccgaccag	gagatgtgac	acagactgta	ccgacagatg	cacctgcaga	15300
gctgttcctt	ttgtcaagtc	cccagggtcc	ggtaggagtt	ctcttcgatc	agcgaggcac	15360
15atacaccaca	gccccaatgg	tgcccactct	accattccag	agcttctcga	gtcaattttgc	15420
acagaacaga	cagctcattg	ctggtccttg	gcagcaaatg	gcacagggga	caaaccacct	15480
gcataatcaa	gtatctaaca	tgcagccaac	accaataggg	cagccagtag	ctgttgagaca	15540
ggctcaagat	cataaccgag	gatatgatca	gaatcagaat	cagaatcaga	atcaaaacca	15600
gaaccagaat	gataatcaga	atggagtgcg	gccagaagaa	aatgatcgga	tggccaatat	15660
20cgccggacat	ttgtggctga	tcttcaagct	cgctgtcttc	gtctacgtct	tcgctggagg	15720
tgggtgtatt	tacaggcctg	taatgctagg	tgctattgct	gggattgtct	atctggcaca	15780
gatcggcatg	tttgaggatc	agatcaacta	cgtgcgtcgc	cattttgagg	ctcttcttcc	15840
tggtggcgct	atggccgaac	gcgctgcaca	acccatcaac	cagcgcccac	gaggtaacat	15900
atcgcccgag	gaagcagcaa	ggcgaatact	acaacaaaga	caagaacaaa	ggttcgctctg	15960
25gttacgcgag	agcttgctg	gagtcgagcg	cgctttcact	ctcttcattg	ccagtctatt	16020
ccctggtgta	ggcgagagaa	tggttcacgc	acaggaagag	agagagagac	tggagagggt	16080
agcagcacgg	gaagagagag	agagacagga	ggaggaagcg	aggaagcgag	aagaagacgc	16140
cagggcacag	cagcaacagc	agaccgatga	gaaagctagt	gaagccaggg	ttgagatgga	16200
cagtgaagg	actccaagca	gcagttcaaa	gggcaaggag	agggctgagg	agcaacacgt	16260
30tgatgggtca	gcctcatctt	catgaggtgt	cgagaggtat	actctctttc	atacatgttt	16320
atagggtttt	tggttccggt	catcacatgg	catccttctg	tacacattgc	gaggcgagca	16380
gagtcogtta	gttatgagcg	gcattctttg	acatgccctg	ccaaggaatt	tcatcaagta	16440
tttaccaggt	acataccgga	agctagacat	gacatgatgc	actaaacaag	ccttttttgca	16500
tcactaatte	ccctcccatc	taccgactca	tccgttcac	aagtctttta	gccaaaaacg	16560
35ctcttttctaa	atccctgatg	gaaaacaacc	ccagtcagca	ccaggatatcg	taccaccacc	16620
ttgctogtct	gacaaatgac	gacatgagcc	gtaaaacaacg	ggttcattgc	tgcatagtat	16680
ctgtctcttc	tctgggaccg	atctgtaaaa	agcaaaacag	acacgtgtgc	cgcagtggca	16740
gtgcaagggg	ctaagctagc	acgtgcgtcc	attgaaccat	gatttgtcca	gctgcacgct	16800
tgcacatgg	atcggtatcg	agctagacgg	cgggtgtacc	tgcgacaaga	gtgcacctga	16860
40gacagaagca	aaaagcaaaa	anaagagggt	gacgacgact	ggcgggacac	gggacgggac	16920

gggcaagcta	acgacggtgc	aacagagcga	cgtcagtgaa	caatatgcta	ggtgacacat	16980
tatttatgtg	agatagtgtt	ggagagaaga	gtatcatcta	ttgattgaaa	gcattatcat	17040
tactggccaa	gcgtaggagac	gacgatgcaa	gaaacgacag	cgacgatacg	aacatctccc	17100
tcaataaaaa	tgacaagaac	aaggggatat	cgaatgcgac	cctcgggaaa	gatccctcgg	17160
5taaaaacctg	agaaatggag	aacaattaac	caccaggca	agaaaaaaaa	aagtggacat	17220
gtaggttgaa	ttgttttctt	ttgcattttc	ttttgttttag	ttgcgacgac	gtagaccaga	17280
gtcaccgggg	aacaagatgc	cattgggatg	gtacttgtag	cgccatgact	gccagttagc	17340
acatgtcctg	tgcattctctg	cgtatccatc	ttgggccttg	atttcggatt	cgcgctcgag	17400
gaactggaca	ccgagaccgg	cagcaacgtg	gaaggggatg	tggcagtgca	taagccatgc	17460
10gccagggttg	tccgactcga	aagcaaggac	caagtagcct	ccagcgggaa	gatcggccgt	17520
gtcccagacg	atgggggttg	ccgtcttcag	ggttgaaata	tctccgttcc	agactgcgtt	17580
ttcgacctgt	gcgaggacgt	agaagtcgtg	gccgtggagg	tggatggggg	gaggaagtgg	17640
tgggttagaa	ctgttttgtt	ggatgaccca	atattgccac	tgaagccttt	gttaatgacc	17700
attaaacagt	actaagtaca	ggggacgact	tacttgggtg	ttctcgtcga	ctgcaaacac	17760
15gtggcggttg	tttccgtagg	taacattgcc	atccaacacc	gactgcagag	tagggacttc	17820
aagatcaact	gcatgggat	taccgttgac	gagccattgg	accagaccct	gattttgcgt	17880
cacgtcacta	gtccagttag	ggttgaagcc	cacgctcaac	tgttcgggca	tctcctgagg	17940
aacagtcgtc	ttggcatagg	gtacaacatc	ctcatcgtag	cagcccagcg	gaagcgaacc	18000
agtcgtgtct	gggtcttcag	ttggagcgcc	agcatatcgg	aagatactcc	tgatatttgc	18060
20tgcatgggca	ttgggaccgt	cgcagttacc	gccggtacca	acacgtagcc	agtagttgcc	18120
cacagcttca	gttgcggtga	tgatgacttc	ataccgttga	cctggcgacg	ttgttagtga	18180
aacgttcata	aacagtggct	atttcgacta	cattctcgtg	gccgggggtg	cgtcacacag	18240
aacgatgttg	tgataaacia	aacgggcttg	catctggaca	caacttcggg	tcaacttacc	18300
gactgcaagg	accaagctgt	ccgtgtagaa	aggttcaatg	ggcgtgaaat	cagccgaaat	18360
25gacctggaac	tgatgcccat	cgaggccgac	atgaaggtag	ttgttaatac	caacgttcat	18420
caaacgcagc	aagtgagatt	ttcccggagt	taggatcggt	tcggcgctact	tgccgccaaa	18480
agatgaggtc	atggagccat	tgacaaggac	attgtcagca	gttggagggc	catttgcattg	18540
aacggctgca	gcgttgacgg	tgaagggtgt	tgcgtgaaac	cagtcagtca	ttgggaaagc	18600
gccaagatca	atatcgtagt	tcgccgttga	gggtcctttg	atgatcagag	gaccacgat	18660
30gccgtcacca	tactgcaccg	agtagtgcca	gtgataccac	tgtagcgggt	agccgtcttc	18720
ggaaacatgt	catacacaca	cagtgtgtga	tacttacggg	agtgccatat	tgagttgctt	18780
tgaatctgta	gagcttgagg	tcaccgggtg	cgattgggca	ttcagtgata	ccatttacgc	18840
catcttggtc	gtttgtcccg	agttgcctca	gacggtgcca	atgtatacct	gtaccgttgt	18900
tttcaaggcc	attggtaact	gtgatctcta	gaacatctcc	ccagtcggca	gtaatagtct	18960
35atgactcatt	agaaaagaat	tattggcatt	ttaacacagc	acttactggg	cctgggtatt	19020
ggccattaat	caagaacatc	ggcctttcaa	aaccgtctgg	agctccagtg	gtattgggtga	19080
tggtcagggtg	atacttgact	gtctttccag	tatctggcca	ttcaacatcc	atgtcgggtg	19140
cgatgttgaa	gtcgtcaatc	cagcatcctc	ttgattctgg	accatgatta	caggcagttc	19200
catatccttg	tcgtttgctg	tgtccactga	agagaccagt	gttcttccaa	ggaacctccg	19260
40gtgttctggg	ggcgagactt	ttatgaggta	cacagatgaa	gtgacagtg	gcaaaagaag	19320

cccaagggcg	gtaaccaccc	tcgaaattga	agagaccatg	atgtagtaat	gaaagctcaa	19380
gaaaaagacg	caattgcaaa	ggaaagctac	cggacccttc	aaaagtaggg	ccaacgaatg	19440
cagctccgac	caagattctg	atgtacaaag	agaagcgaag	gctggctggc	ctgccagaga	19500
atgggtggta	caagtaatat	aaaccaaagc	atggcagact	ggccttgatt	cattagtggg	19560
5tcgaacaata	gccctaaacg	gcacgggtga	ctagtgggtga	agtgttccca	ccactagtaa	19620
ttgacatgat	acctccatat	caaagtgtag	ccgcggcttg	tggcagatac	caagatcgcc	19680
atcacgacgg	tgtgcggcac	atgaacggta	tagatgctga	ttatgaccac	ttgcccggaa	19740
aggcgagcac	aagagacggc	tgcatatcag	ccctaagcag	aaaaaaggag	aattttaagt	19800
ggttgaggag	gaggtgaatg	tcggtattac	ttgcgttgca	cgaccgaatg	cagctagaca	19860
10taccgaagct	ggctccaaac	ctcgaccgag	gacgacagac	cagacggtcg	gcagagctgc	19920
aggcatgtct	gagctgtgta	ctttgggaaa	gccacttcag	caggctgtgg	agcggctgca	19980
gggggtagtg	gtgccttggc	tacgcatgaa	gcgggggtctc	agaggactag	tttgacatgt	20040
cgggtgcaagg	cgtagatggg	atactatgga	tagcacgcgg	gccagggccc	ggaaccgggc	20100
tctccatgaa	gttgagaagc	gtctccgagg	cttggaagc	ttgatagtcg	agctagcgag	20160
15tagaagaatc	tcgagaaggg	caagcgagtc	gcgacgatct	gtactaatgt	gataagcgct	20220
aggccgcgtc	cggcgagcgt	ctcggacttc	tgagaggggt	cccgaacctc	tggtccgacg	20280
acaacatcac	catctcgtaa	ccctggcctc	tgagagcgca	tcgctgtccg	tcttagggcc	20340
aatctggcat	tttcaactcg	atacaatcgc	ctgattggaa	tggtttccac	tggtttccatg	20400
acgtttcgtc	ccaaacgaag	actattgata	tgcattccgaa	tgcaccggct	gccttagcta	20460
20ggcacaatat	tacagctcaa	ttgaggcctc	gatgtttcgg	tacagcttgc	aatgctcaca	20520
tcttgccact	ttaaagtgcc	tgagtcgaag	gcgcgtgct	tcgctacgtg	gcttggggag	20580
caatgttggc	atccgacagc	ctgcaggcga	cgggaaagat	tttgccaata	cagagtgtga	20640
caatcatgga	cttgagtagg	ctttgcaaca	catgcaaaact	ggggtaacag	tggagaatcg	20700
cagggtcgca	agaccctttg	cgacgccacc	ctaccgggtga	ggcttcacgg	gcgtttgctc	20760
25acagcattac	tcgcgccatg	cagaacgact	tcgcagaaaa	gaaatgctca	acagtattcc	20820
ctaagatctg	atactgctgt	accgaaacgc	cttagtgggt	acgggcatga	cagcgaagcg	20880
cgccacaata	ttggagaggc	agtgtgcatg	tctgcttact	tgcaggctac	agcaaagctc	20940
catactcagc	tcgccgtggc	tcatattact	ccaccagatg	caagacacca	acttgctca	21000
tgctcaccgt	gttgttgctg	aaggcggagt	aggatccaga	cacacgccc	ttgacgtaac	21060
30agttagctgt	aaggcgtcta	tttccggtac	acagtcagcg	ctacacgttc	aaagaagaaa	21120
cgagcatcag	naaaacatct	tgtcaagtcc	tcggctcgag	tcagactaga	gaaaagtaac	21180
ttgacatcgc	cggccaattg	gcgccgaaat	taaaaagaaa	caccattgta	cagttggggc	21240
ccaggctgca	gcagtttaat	gtcgtactg	caaagtacct	ggttcgagta	tcctgcgcat	21300
ctgcaccatc	gctcgaccct	ggtcgccgtt	tattcagata	aaagctccgg	gactaagatg	21360
35tagtcgcatg	gttgtgcata	ctaaactggg	ccgatcaagg	gacgccaaca	cgcttgtctg	21420
ctgatcgatt	gccgttatcc	gtacaaccaa	agacacagga	aaagagccgc	ctgaatggac	21480
cgagaaaactt	cctgatgttt	tcagcgtttt	aacagatcta	ccaggcacac	cggatcaacc	21540
tggattatct	ttgacagtac	ttggtcatta	tcgcgttata	agtggaataa	aagtatgtac	21600
aagaccagag	cagatactca	cggtaggaac	acaggtttct	cagcatccat	ataccttgtt	21660
40gtatcgtcac	acatgttgat	catctcctct	gcaagtaatc	cactccaaca	tcccataggt	21720

caatagcaag	atgtaagtga	ttgaaactct	cactgntcct	gcatcatgtg	ctacctacgg	21780
gctcttccgt	accagcaatc	tctcgagcaa	gcaatcttgc	ttccgagatc	ttaggcaggg	21840
tatctcgcaca	agcgaatata	tatgtattga	tgacgaaacc	cccatgtctg	gtctctgaga	21900
gggctatgtg	caaatagcct	gaatgatcct	acgtctgccg	ggggatctac	ggcaagcaaa	21960
5gtgtttttct	agacgagtcg	aagagaaaag	agtagaggag	aagatgttta	caattcctag	22020
gtggatggga	gtaccgaatt	cgtttggtgt	tacgctcatg	ttgagcaaca	acagctgtgt	22080
cactcgctcc	actcgttgaa	atctcgtata	tgcaacggat	gcggataacg	tagattgaat	22140
gatggtacac	tggtaaccct	ggtgtatcgc	aagtaagtga	ccctctcttt	ctctgtagt	22200
gttccttgca	gccatcaaga	catggtcttc	gccacgtgcg	cacatccaca	gtgctcgacg	22260
10cgcggtcgc	gtggaagccc	catagattct	ctagatcgtc	aatcgatgtg	gcgcatgtat	22320
cagcacgttt	catacattga	acgcgcaccc	cgtaccagaa	gtaaaaatag	taaagcttaa	22380
ttctgagcgt	agcagatgat	gctcagccta	cgccatgaca	gatggatggc	ttgactcgag	22440
cacggatgta	tactactcga	ctagccccc	cggctactgg	ctatggctaa	atgtgaatgt	22500
cgtacgcata	catgctatgg	ccgcttcagt	gcatgtgtta	acttaggggg	ccaaacgata	22560
15tcagcctgaa	cgtggggcaa	tctattttct	tcccttgga	tggagtgacc	tcgcataaga	22620
cgtacatat	gtaactcagc	ttagacataa	ccatgctttt	ctccacaaaa	ggtctgcaca	22680
gatatcttgc	tgaatgctta	acgaacctcc	ttaatgccag	aaataaaccc	gaaggtgtcg	22740
tcgcctttcg	cactcttatt	ccaccaaata	cacactacag	ctgtcaaagc	aacactacaa	22800
atacacaacc	ggagggtctc	gatgccacaca	gtacctcat	tctcgggttg	aattacgact	22860
20ttgaaacgta	cccttttagga	ttcaccaaaa	taacacaaaag	agacccaaca	ccctcgacta	22920
gacgataaca	ttttgtaccc	ctcgtacccg	cagcatctat	cccttacttc	tcttctccgt	22980
accctgactc	caggttctgc	actaccttca	tgaacaaaac	agcaggagag	agaagctcga	23040
aaagcactcc	cagttgacct	aaaattcttt	gaacttacaa	atgcgacatg	gctcagtgg	23100
caatagtagg	aatggtgaaa	aatgcgcggt	gtctggagct	agaccgggaa	tcataacctt	23160
25gctagaagat	ttaccgaagc	agtattggtt	gtatagtgg	aaggataaaa	cgctactgtg	23220
tatgaagtta	gcatcggtgc	catggccttt	acaacaagac	cagcgaacat	tcatatcatt	23280
tcttgatttg	gctgaaaaat	accaagtcgt	tgcgtaacac	attataccaa	aactccacat	23340
atatcatcat	gctaataata	acccaaccca	gatccaaaat	cccttacatc	accatcttaa	23400
acgccggatc	cctatcatct	cctcgtcctt	cgaattccgc	atactttcca	ctcgcccttat	23460
30cgcggttgat	tttccacacc	cacgccacat	tcgctgcgac	cctatcattc	acgttagctg	23520
catatcccgc	cagcaaaacg	atggaaatcc	acctgtacac	actagaaaag	agaacaagga	23580
gaagagaact	tacaaaaccc	aagccgcaca	caacaacccc	agcataatgc	tatgactctt	23640
atggaaattc	ggcttctcgc	tcttctgata	cgtaaagctc	gcgacgaacc	aaccgcatt	23700
ggcaatcgca	agctgcagcg	ccgatgtagt	cgcgcgcttg	tagtggccag	cggaattatt	23760
35gctgttccaa	gaaagaatac	atgggacgga	ggaatacatg	cctgtagcca	tgagaaatgt	23820
cattccgtat	tgtactttcg	ggttcgtcga	ttggctgatg	actccgtagc	cgatgatagc	23880
aacgggaagg	gtacacagca	tgactgggcc	acgtagagcg	aggcggctcg	agagaatggc	23940
tacaaggacg	gtgaagacgg	aagcgacggc	gtaaggaatg	acggtcacaga	gctgggcttt	24000
gttgggggtcc	ttggcgaaagc	cgttgttgat	gattgtgggg	aggaagaggc	cgaaggagta	24060
40gaggcctgag	aggatagaga	agtaggcagt	ggctgtgagc	cagacttgga	ggttgaagat	24120

gccgcgtttg	atttcggccc	agtcgaagcg	ttcgtgcgag	gcggatgctt	caccgagtcg	24180
gaagcgggcg	tgggcttttt	cggaggggct	aaggaaaccg	gctgattcga	tgggaattggg	24240
aaggaagata	gcagagcatg	cgccgacgcc	aacggttatc	aagccctcaa	ctatcaggat	24300
ccatctccag	ccttcgagtc	cgcttgctgg	gccaatggca	ttgaggcctc	gagcgaggag	24360
5tccgcacaaa	gcaccagata	gagaggctgc	agtatagaag	atgcctatgc	gtagagcgag	24420
ctcctggcgg	cgataaaagt	gagagagata	tagtcctagc	aaggcgtag	cgaaagacgg	24480
tttattacgt	cgccaaaatc	ttaccattcc	aggcaatagg	cctccttcag	caacgcccag	24540
gagcgcgcga	acagaagcaa	aagacgcgaa	atttgtcaca	aatcccaagc	acatggctag	24600
gacgccccag	atggctgtca	gaaagggtaa	ccatatcttt	ggcgatacct	ttttcagcag	24660
10caaatggac	gggagttcgc	tgcacaaatt	tagtcttctg	gactgcatct	gccatacagg	24720
aagcttacct	cgcaatatac	gtagcgtaaa	agacgcaaag	cccaatagcg	tactgggtgt	24780
ccgtaagatg	gagatcattc	tccaaaccaa	gaatcttcgc	attcccaagg	tttgtacgat	24840
caatgaacga	gcacaggaac	agcagggcga	gaaccggcag	gatgctagaa	aacaacaaaa	24900
ataaacgagt	cagccacgta	catcatccac	aaaagccctc	aagataaaga	catggagagc	24960
15aggaatggag	agagaaacat	accgacaatc	gatcttgaac	aacaacctac	tcgttgctct	25020
cggatcctcc	agaagcccct	cttcaaacc	ctcactccca	ctccgctcaa	ccttttcatc	25080
ctgcatttta	tccacgctca	tcgtactcgc	cctctatctc	gcaccgtctc	ttatctccaa	25140
ccctttcgcc	tctctgagct	atctcaaagg	ccccaaagta	ataaaaaaaaa	gagacagagc	25200
caaagaaagc	aggtagaaag	actgagtcgg	ctgccacgag	caacgtagca	agtaagcagc	25260
20gcaacacgct	aacatccaag	cacaggaaca	cacccctccc	catcctttta	aacatcccca	25320
acccccctc	tctcccctcc	atctcagcta	aacccctgca	gcgctaccca	tcccgacccc	25380
acggggcatt	ctccaccgca	actcagcact	cagcacacag	caacgggttg	acggctccac	25440
cacgtatccc	cccttgcgca	cttgattggc	gcagcacgcc	gctcggaacn	caatagtagc	25500
ccacatgctg	gcccggcttg	tgcgttagcg	gtaaagaagc	agcaaagcga	tcagtccggc	25560
25gctgtgccac	gcttgacggg	cctttttttg	cggccgaaag	aagggtgctc	aaagcaaaaa	25620
aaacacacat	gacaaggggtg	tgagtgtgtg	tgtgttttagg	gttgctttgt	caagaacatc	25680
atttttacgt	atgtctgcgg	tcaagcaaga	ggaatttgcg	ttgcacatga	gcgatgggtg	25740
gcgtcccttt	tgggagcgaa	ggagcgagcg	aagattgcat	gggcgtcctg	tgtaccaagc	25800
tttttttttc	gtgacaggcg	tggcgaacca	agaggtgcga	agccttttcg	cttgctgtgt	25860
30gcttggtgtg	gcttggtctaa	catttccttg	gttcgcggcg	cgtcaacttg	agtttttgtg	25920
gggggggtgt	tggcgctacg	tgtgtcagcc	aggaattttg	aagaggcttt	ttgcacatac	25980
acatacacat	acacacacac	ctcatcaagc	ggagcaaagt	agatagtgac	agaggactta	26040
ctttttttta	ttgccatgtt	tcccattttg	gaaggaggaa	aaaggtagat	aagtgactta	26100
cttacgcgct	ggaagatgca	tgcgttgccg	gcttntagac	cgtcctttta	gcaccttgct	26160
35agataagaaa	aaggtgggga	cggaaagtat	acccggtaca	ccgcaactatg	tacaaagagc	26220
tactactgca	acatagagaa	acaaaaatgc	cactactact	ccatcgtctt	gagatcactc	26280
gcgacttcaa	acttcgcatc	cgtcttcgcc	ttgacatcct	caacgctaac	ccccggcgcc	26340
gtctccgtca	gcgtcaaagt	ccccctcttc	ctgtttactt	caaagacaca	cagatcggta	26400
ataatagtgc	tcacgcactt	tgctcctgta	agcggcaact	ggcattcctg	aacaatcttg	26460
40ctggatccat	cttttagcaac	gtgttcagtc	gcgacaacga	cttttgtagc	atcgggattg	26520

	ctaacgagat	ccatggcgcc	gcccataccc	ttgaagactt	tgccgggcac	catgtagttg	26580
	gccaggtcgc	cagaggcact	gacttgtaga	gctccaagga	tggatacgtc	gacgtggccg	26640
	ccgcggatca	tgccaaagga	ttcggcgctg	tcaaaggctg	aggcgccggg	aaggaggggt	26700
	acggttttctt	tgccggcggt	gacaatgtct	gcgtctactt	cttcttccgt	cgggtagggg	26760
5	cccatctcta	gaatgccatt	ttcggattgc	agccacacct	tgacgccatc	gggtacgaat	26820
	gctgctgcgg	ctgtggggat	gccgacgccc	agattgacgt	agtatccctg	cttgagctcc	26880
	tttgctgccc	gacgagcgat	acgatcgctg	cgttcggcgg	cttcgttctt	tgatgaggcg	26940
	tctttggatg	cagccggttt	tcgcagcttc	ttgatctcaa	tgttcttggg	ggcggtaggt	27000
	gggacgatgc	ggtcgacgaa	gatgccaggg	agatcgacct	cgttggcatc	aaagggtgcct	27060
10	atagggacaa	tctcttcggc	ttcgacaatt	gtaaggcgtg	cggctttggc	catgatgggt	27120
	ccaaaagctt	tgggtgggtg	tctgctcata	ttttcagtga	agctcaaggg	attgggatct	27180
	ctcaattcnt	acctgaaaac	acagttacca	gcttcatcgg	ccttggtgtg	acggataatg	27240
	gcgacatcgc	cgggtcaatgc	agtctccatg	aggaaacttct	tgccattgaa	ctctctaacc	27300
	tcacgcttct	gtccgtagcc	tacagccttg	ccctccttgt	caaacttggc	cggaatctgg	27360
15	ccatcttgca	acagtgtatt	tactgcagtg	gggtgtgtaa	atgctgggat	gcctgcgcca	27420
	ccagcgcgta	tcctctctgc	aagcgtacct	tgcggacaaa	gctcaatttc	aataccaccg	27480
	ctcagatact	gcttctcaag	cgccttggtg	ttgccgagaa	aacttataat	gagcttcttg	27540
	acttgctcgt	tctttgtaag	atgtgccaat	cctcctacgt	cttcaatgcc	agcattgttt	27600
	gagacggctg	ttaacgaatg	taatgactcc	ggcccacgct	tcttcatcgc	tgcatcaaaa	27660
20	gtgtctgcga	caccacacaa	cccgaatcct	gcgctcagga	cggtggaacc	aggctgtaca	27720
	tctgcaactg	cttcgtctgc	atctttgaaa	agctttgatt	tcgagcggtc	aattgtcggc	27780
	gcgcgctcac	tgatgcagcg	caattgacct	gtaagccgcc	atcgtggtgg	cagagttcgc	27840
	gcagcccgc	attgtggagg	aagagatcgc	cgggcgcata	gccgtgaggg	aagagctgtg	27900
	agcagccggc	aggaagcagg	cagggtatcc	attgtgaagt	aaattacgga	cgcagcaaca	27960
25	gcgtgaacgg	tctgggttaa	tcatttgaga	aggggtttca	acaaatgccg	actcatccaa	28020
	gcgggcgacg	atttcgcggg	cgagctccga	acatgggtcc	tggagcgcgt	gcgacaatgg	28080
	cactgcaact	ctaacgtcat	gcattctttg	atccgtcggg	gccagttcaa	gtgccgaacg	28140
	gcgagcgacc	ttggagcttg	gagcggggct	tcgtctaccg	cctcggtcgg	atcaggctta	28200
	ggcgcgctgc	cgtcctcctg	gcgaagcccg	aggttcatcc	accctctgca	tccacacgct	28260
30	tggccacttg	ctagtccac	gagcaagacc	ggcgtcccg	taatacgggg	agtcgatacg	28320
	catctctgcc	gtgccatcag	ggaaagttga	agtcagggaag	attcatccag	tctgtccata	28380
	gcgggttgga	gtcgttccag	ggcagatcga	gcaatgctgt	agagaagaga	tccgagtttg	28440
	taacagtctc	attgacctcg	ttataggaca	ttgcagagaa	atctaacctg	tcaggccagt	28500
	aatcgcttac	tggatcatgt	cccaagggtc	gactattagg	atttggcatc	tcattgtgca	28560
35	tcagtctcga	gggcgaaagg	agatctggta	tagccattga	ggttgggtgca	ggctcctgcg	28620
	gaggtagggg	cgggtggcatc	tcaatgtcag	gcgcctgctt	aaccaatacg	cgcagaaatc	28680
	taccatacac	aactgatgcc	ccgttccgat	ggacagggtg	actgccaatg	cgttccagga	28740
	cggttgccgt	atcctctatc	aggtgccgca	cgttgggggc	aaggctcttc	ttattgccac	28800
	tcccctcggg	tacgggtgcg	cttagagcca	gcgctatgct	ggcttgcaaa	gcaaatcatc	28860
40	gtgacgggtg	tattgggcat	tgacttgaga	ggtccctcgc	cttggattgc	tgccggcgcat	28920

	aacattgagc	gccgaggata	acgcagactt	gcggaacgag	cgcttgactt	ctggtggcgc	28980
	gctcggatgg	ttcagaagca	ttgagtaggt	cgagagtcgt	gtgtgtgtaa	cgagtatctc	29040
	gacataaggg	ggtaggctct	tgtctctctg	gtcacttata	actgaaggcc	atgcccagaga	29100
	ccaattgtcg	aagaagccct	caattgactt	gtcgatttcc	tgcgcaacct	gggaacctac	29160
5	ttcggccgag	ccatagttgt	cgcatcgctg	gcgtacttcg	gcaaaaaggt	tgtcgagatc	29220
	gcgacgtaat	actgccatgg	atactagcgg	accgtcctgg	gcatccgagt	gctggtggctc	29280
	atgccattta	tcgctgtatt	gaatcaagca	cgtcttttgg	acacagtagc	tgcggccacg	29340
	agcgaggcac	acgccacgct	ctagcacaaa	cagcgcaatc	cagacccttt	ctctccgacg	29400
	aagcagtcgc	tggccccatt	cagaagtcgg	gtcaatgtcc	tcgaaacat	ccatagctag	29460
10	tgccttttctt	gcgtcaagac	actctgcttt	gggcatctgc	ctcgtgagct	ccggacccaa	29520
	ggacgtagat	ggagtgatga	ctttgtctag	cataagatcc	aaagaaatag	acaatgccgt	29580
	agctagatag	agacttggtg	cgctcgctgc	tgcattgcgc	cctgggggca	tccatggtat	29640
	gctcaccatg	aatgccagga	cgatttcaac	ggatctgtac	tttcgaacaa	tgacctgttc	29700
	agctagaaac	ctgcggtgaa	gaagtagtct	tttggccaaa	gcagacgttt	ctggtaggaa	29760
15	gacggccgctc	acagccagca	atgtcgtaaa	cagaaaggct	gagcgggttc	ggacaaaagg	29820
	tagagtatgc	accactgggt	ctagacccca	gcgcgtgtga	gctagtcttt	tgtggaaact	29880
	gcgtcttgctc	agctacttga	tggaaattct	ttggcaatac	gtactattgg	aggagcatct	29940
	ccgcttcac	tttggttaacc	aatcctacat	caattggatc	taaccagat	ccttggtcca	30000
	aatgcgcctt	cattggtaag	aagaagtgat	gaacatcgag	gaatgcgcct	tgctcgctac	30060
20	cagtaaacct	gccttctggg	ctggcgacac	ttgtattgta	cgactgtggg	gtggtggcaa	30120
	tcctcaagtc	tgatgcgcgg	gctaaaagct	gaagcggatt	ctcgacatct	tcaacggcaa	30180
	gctgatcatc	gcttgaagtg	ctggcaactt	cttttgctgg	cacataagat	agttctgcta	30240
	gtactggcgg	tgattttgca	tcttgactag	ggccaacgct	tccttgctgct	tcgttcaaaa	30300
	gctgttgcaa	atgctgtaac	gtgctctggg	tgacagctac	gtctgatttt	ctcttcttga	30360
25	ttgcttcttc	tacttggtag	attgctttct	ccaaccctga	tcgtttgcta	gaacttgtaa	30420
	gcttagctaa	tcacactagg	aaacgatact	cactttttca	cgcccttttg	cctaccacta	30480
	attgccagta	catcaagtca	gcgcattgtt	cgctatgggt	ttgtatcggg	acgtgacgta	30540
	catatggaac	tcgggtataa	tgcatcgcac	gcctacgcta	gagcatcttt	cacacacaga	30600
	agcgccttct	tcgcgcggc	attttatctt	gcttttgctg	caattcagac	atgccgcttg	30660
30	cttgacgttc	atgtcttggg	tggcttcttg	gcgggctgag	tgttaggaga	tttggttagg	30720
	gtgcgccgca	gatgacagca	acatggtaga	cagtcggcaa	tgttggccaa	gtcagatcta	30780
	gatctcagca	aaggagctag	gagcgacttg	cttggatggt	ggaggtacac	tgccgacggg	30840
	agcttcggag	aatccaagtg	tgagggccat	gctagcccga	gaccggcatt	gcgctaattg	30900
	gaccctggcc	tgtaacgtgg	gaaggacgaa	cagcacagg	gcaggcttct	agggctgcat	30960
35	gcagtgcgca	tcactctgat	gcacttgctg	tgccaagtcg	tgtactacac	aagtgcgagt	31020
	tgctatttgt	aacgaggaac	cttgtattta	aaagtgtata	cgtgaggtag	gtgtgttcca	31080
	gacctccaaa	tctaaagcta	ctaaaacaat	agaaacagcg	gagtctactc	cgacaaggctc	31140
	aagtgaaagg	cggcggcata	aaagtcaatc	gaatcaaagt	acacggacat	acgagcaatc	31200
	tacacacggg	catggctata	gcttactttc	gttctgcttc	aatcgtatga	cgccttatctc	31260
40	atgtaagcac	agtctactat	agcagacata	agcaagctgc	ttacctcttg	gacgcagctg	31320

gcaatgagcg	tgccatcctt	ggtatacatt	ctctgggaaa	cgaggccgcg	accatcacca	31380
gccccagggg	tctccatctc	ggtgaagatc	cattcatctg	cgcggaaact	gcgaggattg	31440
tgaaagtaga	tggtgtggtc	cagactaacc	atcatgccaa	tctcaggctt	tgcgctctccg	31500
ctcttttgcca	ggtcttctgc	tttcctcaat	tcacgtatgc	gctgcttgtc	tgattcgttg	31560
5acaaaagcttt	ggcgctgtaa	ctcggcatac	tccatctcga	gcagcttctt	aagtacgtcc	31620
tcgtcgatgc	tcgacctggc	cctgctcttg	cgctggttcg	agtagcgag	aagcttggtc	31680
gcacgcgcga	cggtgccgat	gaagtagcta	tcggacatgt	atgcgatggc	ggagagatgg	31740
gcttcgtgac	cgcagcggg	ggagatttta	cgcgcagcct	ttatccattg	tcggcatttc	31800
ttggtgtggg	gcttgctcga	gtcgtctgct	agatatgagc	tagggcaggc	ttaaggatgg	31860
10tatgcgaagt	cactcacctg	tttcaatggg	caacagctgg	gtctggaagg	gactctggcc	31920
atcggtgggc	gtcttcaagt	cgctcgtacc	ttccttgggc	gccgggacgt	ctggcatcgg	31980
gtagatgtgc	tcgacctttt	gagcgcctcc	actgttctgg	cgaacaaaac	tcattggtcgt	32040
agtgaagatg	acgttgcccc	tttgccgggc	ctgcaccgtc	ctggttgcga	acgactttcc	32100
cgagcgcacc	ctttctacat	ggtatatgac	ggggatctcg	gagttgcctg	caaggatgaa	32160
15gtagcagtgc	atcgaatgca	cagtgaagtc	ggggtcaacc	gtcttctggg	cggcgctgag	32220
tgtctgggca	atggcagcac	cgccaaagat	gccgcgcgca	cgggggggat	gccatagggg	32280
acgagtgttt	gtgaagatgt	tgggatcaat	gtcggccagc	tgcgtcagtt	caaggacgtt	32340
ctcaatggcc	gactgggagt	ggtcggcggg	cggggggcgg	atgaggggtg	ccatgggtgt	32400
ggctgatagt	tttcctgttg	gtggatcggt	ctgtgttctg	cgaaaaggag	gccagtgtag	32460
20caagaccaga	tgcaagcagc	agcagcagc	ggctgtgtga	gactttgggc	gtcgtcattt	32520
cgggggcacg	tcaaagcagc	gcagacgcgc	atgagccgag	gcacaatgat	catcgcccat	32580
gtgggagctt	gtcgcgccga	acacgtgact	ggccgctgac	tgatgggggc	tgactaagcc	32640
aggcggcgcc	aagccgagga	gcaggctggc	tctggggtaa	aaacgtcata	ctgggcttgc	32700
cgggccctgc	gcagatgcgt	acctggcttg	gtgccagaag	actaccact	cttgcatacc	32760
25tacatagtca	atgtttcatc	tgtcacctgc	tgtccgtcct	cgacgcgtgc	ccgcntctgc	32820
atgentggtg	cattgttctc	actccttccg	ctaggcgcc	tgtatctgca	tcttcctgt	32880
gcctgcgcct	gtgcttgtgc	ctgtggaatg	tcgcggcccg	ctgctgcata	gcctatctgt	32940
acatacaaca	ccatcccatc	ccgcttcacc	tgccttgcc	ccctcctcgt	gccacacatc	33000
cgcgcgccac	aacaccatgg	ctgcgaccaa	ccccgagctg	caggccaaac	tgaggagct	33060
30ggaccacgag	ctcgaggagg	gcgatattac	acaaaaagg	tccgtactgc	tgaccacca	33120
cgcctatccg	cctctctgcg	tgcgctaata	agtcgcatag	ctatgaaaaa	cgtcgcaccg	33180
tgtctgtgtc	gcagtatcta	gggcctgact	ttgctgccc	gttgccaggcc	gacctgaacc	33240
agcagaaccc	acccaacca	tccagtgagg	gctctcgctc	ccgcaccgca	tcctttgcta	33300
ttccgtccgg	tccgagtcca	tcacngcgac	cacaaccccc	acatatccag	ctccccgc	33360
35ccgactcata	ccatgacgct	tccgcacagg	gccaattggg	cgcacccatg	ccatatgcga	33420
acgcctccgc	cgctgcctcg	gggggctcgc	agtacatggc	atacccgccc	agccaagtgc	33480
gccgttttca	agagaagcag	ctgggcctgc	gtacaaattc	gctccagcgc	aattcctcac	33540
agctgtcgca	aggaagcgag	acgttcattc	cacggcctca	aacgcctgaa	tacaaccact	33600
cgcgcgagcc	caccatgatg	ggcaactacg	ccttcaatcc	agacaatcag	caaagtattg	33660
40atggccaatt	tggctctccg	ggagaggcca	gtcgaaggag	caccatgctc	gaggtaaacc	33720

	aggggtatttt	ttccgacttc	acaggccagc	agatgcaaga	caatcgcgac	tcgtatgggg	33780
	gaccaaacgg	ctactcgtcg	ggagatgcct	tttctcctac	cgccgcgatt	ccacctccca	33840
	tgatgaaccc	caacgatctc	cccttgggcg	ctgctgaaac	catgatgccg	ctagagcccc	33900
	gcgatctgcc	ttttgacgtt	tacgaccctc	acaaccccaa	tgtcaaaatg	tcaaagtttg	33960
5	acaacattgg	cgctgtcttg	cgtcaccgaa	gtcgcacaca	gccaaaggacg	actgccttct	34020
	gggtccttga	cgcaaaaggc	aaagagacgg	cgtccatcac	ctgggaaaag	gtggctagtc	34080
	gcgcggaaaa	ggtggccaaa	gtgattcggg	acaagagcaa	cctctatcga	ggcgaccgtg	34140
	tggcattagt	gtacagggat	acagaaatca	ttgattttgt	cgtggcggtt	atgggctgct	34200
	tcattgcggg	cgttgtagcg	gtacccatca	atagcgtcga	cgactaccag	aaactcattc	34260
10	ttctcctaac	gacaactcaa	gctcatctcg	cattgaccac	agacaacaat	ctcaaggcct	34320
	ttcatcgtga	cattagtcag	aaccgtctga	aatggccgag	tggggtagag	tgggtggaaga	34380
	cgaacgagtt	tggcagccac	cacccaaga	aacatgacga	tactccagct	ttgcaagtac	34440
	cagaggttgc	ctatattgag	ttctcgcgtg	cacctactgg	tgaccttcgc	ggtgtggtgc	34500
	ttagtcaccg	gactattatg	caccaaattg	cctgcacacg	tgccatgatt	agcacgatac	34560
15	ccaccaacgc	tcagagccaa	gacacgttca	gcactagcct	acgggatgca	gagggaaagt	34620
	tcgttgctcc	agcaccgctc	agaaacccca	cagaagtgat	cctcacgtac	ctcgaccgcg	34680
	gcgaaagcgc	tgggtctcatt	ctcagtgtct	tgtttgcagt	ttatggaggc	cacaccaccg	34740
	tatggctcga	gacagcgacc	atggaaaccc	cgggtctata	tgcacatctc	atcaccaaat	34800
	acaagtccaa	catactgcta	gcggattacc	caggcctcaa	gcgcgctgca	tacaactacc	34860
20	aacaggatcc	aatggctaca	agaaacttca	agaaaaacac	agaacccaac	ttcgcctccg	34920
	tgaagatctg	tctgattgac	acgcttaccg	tgcactgtga	atttcacgaa	attctcggag	34980
	atcgatatatt	caggccactg	cgaaacccta	gagcgcgaga	actgatcgcg	ccaatgctct	35040
	gcttgccaga	acatgggtgga	atgataatat	ctgtacgcga	ctggctaggt	ggagaggagc	35100
	gcatgggctg	cccgtctaagc	atagcagtag	aagagtcaga	taatgatgaa	gatgatacag	35160
25	aggataagta	tgcagcggca	aatggctact	ccagtcttat	tgggtggtggc	actacaaaga	35220
	acaaaaagga	gaagaagaag	aaaggcccgga	cagagcttac	agaaatcttg	ctggacaagg	35280
	aagctctgaa	gatgaacgaa	gtcattgttc	tggccattgg	agaagaagca	agcaagcggg	35340
	caaacgagcc	cggcaccatg	cgagtcggtg	cctttggata	ccccataccg	gatgcgacac	35400
	tagctattgt	agaccctgag	acaagtcttc	tatgttcacc	atactcgata	ggcgagatct	35460
30	gggtagattc	gccttcactc	tctgggtggct	tctggcagct	gcagaagcat	acagagacca	35520
	ttttccatgc	tcgaccatac	cgtttcggtg	agggtagccc	tacgccacag	ttgcttgaac	35580
	tcgagtttct	gcgtactgga	ctcctcggct	ttgttgtaga	gggaaaaata	tttgtccttg	35640
	gactgtacga	agatcgcac	agacagcgtg	ttgaatgggt	agaaaatggg	cagcttgaag	35700
	ccgagcatcg	atactttttt	gtgcagcacc	tggtcacaag	cattatgaag	gccgtgccaa	35760
35	aaattttacga	ctggtaagtg	agctgccaac	agagcaagga	ctgtctaacg	tgtcatagct	35820
	cgtcgtttga	ttcttatgta	aatgggtgaat	acctgccaat	cattctcatc	gagacgcagg	35880
	ccgcatcgac	tgcgcccaca	aaccaggtg	gaccaccaca	acaattggat	ataccatttt	35940
	tggattcact	atctgagagg	tgcatggagg	tcctttacca	agagcatcat	ttacgggtat	36000
	actgcgtgat	gattacagca	cctaatacac	ttccacgagt	catcaagaac	ggacggcgag	36060
40	aaattggcaa	tatgctgtgt	aggagagagt	ttgacaatgg	ctctctgccc	tgtgtacacg	36120

taaagtttgg	cattgagcga	tcagtgcaga	acattgcgct	cggtagcgat	cccgtggcg	36180
gcatgtggtc	atttgaggca	tcaatggcac	gtcagcaatt	cttgatgctc	caagacaagc	36240
aatactctgg	tgtcgatcat	cgcgaagtcg	tcattgacga	caggacatcg	actccactca	36300
atcagttctc	gaatatccac	gacctgatgc	aatggcggtg	atctcggcag	gccgaggaac	36360
5ttgcttactg	cactgtcgac	ggtcgaggaa	aagagggcaa	aggcgtcaat	tggaagaagt	36420
ttgatcaaaa	ggttgcgggc	gtagcaatgt	acctcaagaa	caaggtcaag	gtccaggccg	36480
gcgatcatct	ccttctgatg	tacacgcatt	cagaagaatt	tgtttatgct	gttcatgcat	36540
gttttgtgct	tggagctgtt	tgcataccaa	tggcgccaat	tgatcagaac	cggttgaatg	36600
aggatgcgcc	ggccttgctg	catatccttg	cagatttcaa	ggtcaaagcc	attcctgtca	36660
10acgctgacgt	tgaccatctg	atgaagatca	agcaagtatc	gcagcacatc	aaacaatcgg	36720
ccgctatcct	caagatcagt	gtgccaaaca	catacagcac	aacaaagccg	ccaaagcaat	36780
ccagtggctg	ccgcgacctc	aagcttaca	ttcgaccggc	atggattcag	gcgggtttcc	36840
cagtgcctagt	ctggacatac	tggacgccc	atcaacgtcg	tatcgcagtt	cagctgggcc	36900
atagccaaat	catggcactg	tgaaggtcc	aaaaagaaac	atgccaaatg	acaagtacac	36960
15gaccagtcct	tggttgtgtc	cggagcacga	taggacttgg	tttccttcac	acttgtctca	37020
tgggaatctt	ccttgccgca	cccacatacc	tgggtgtcacc	tgttgacttt	gcacaaaacc	37080
ctaataattct	gttccaaacg	ctttcgcggt	acaagatcaa	ggatgcatac	gcaacgagtc	37140
aaatgttga	ccacgccatc	gcacgcggag	ctggtaagag	tatggctctg	cacgagctga	37200
agaatctcat	gattgcgact	gatggaagac	cacgcgttga	tgtttgtaag	tgaacatttg	37260
20tatgagagga	ctttcatgat	tgctaactca	atgcagacca	aagagtgcgt	gtgcactttg	37320
cgccagccaa	cttagaccca	accgcaatca	acactgtcta	ctcacatgta	ttgaacccaa	37380
tggtagcatc	acgatcatac	atgtgtattg	agccagtcga	gctccatctc	gatgtgcatg	37440
ctctgcgacg	cggcctcgtc	atgcccgttg	accttgacac	agagcccaac	gctttgctcg	37500
tccaagactc	gggcatggtg	ccagtgcgca	cgcaaatatc	cattgtcaac	ccagagacca	37560
25accaactgtg	cttgaacggc	gagtacggcg	agatctgggt	gcagtccgag	gcgaatgctt	37620
atagcttcta	catgtcgaaa	gagcgcttgg	atgcagaacg	cttcaatggg	aggacgattg	37680
acggagaccc	aatgtgcga	tatgttcgta	caggcgattt	aggatttttg	cacagcgtga	37740
cacggcccat	tggacccaac	ggtgcacctg	ttgatatgca	ggtgcttttc	gtgcttgga	37800
gcataggtga	cacttttgaa	gtcaacggac	tgaaccattt	ctctatggac	attgagcagt	37860
30ctgttgaacg	ttgtcaccgg	aatattgtcc	ctggaggctg	gtacgtttct	tcgattcgct	37920
gttatttagt	aaatacttac	taacactcta	cagtgcgtgt	ttccaggcag	gtgggcttgt	37980
tgttgctggt	gtggaaatct	tccgacgcaa	cttcctcgca	agcatggtgc	ctgtgattgt	38040
caatgcaatt	ttgaacgagc	atcagctggt	cattgacatt	gtctcgtttg	tgcaaaaggg	38100
cgacttccac	cggctctgtc	tgggcgagaa	gcaacgcgga	aagattcttg	caggatgggt	38160
35cacacggaag	atgcgcacaa	tagcccagta	cagtatacgg	gatcctaatt	gacaggattc	38220
ccagatgatc	acggaagagc	ctggctccacg	ggctagcatg	actggaagta	tgcttgggcg	38280
aatgggcggc	ccagccagta	tcaaggccgg	gtcgacaaga	gcaccgagtc	taatgggcat	38340
gacagcgact	atgaataatc	tatcccttac	acagcagcaa	cagcagcaat	accaacagcc	38400
gggtatgtat	gctcaacagc	aaggcatgca	ccccagcaa	caacaccaat	ttagcatgtc	38460
40caacacgcca	ccacaaggtc	cacccaagg	cgtagaacta	catgatccta	gcgaccgcac	38520

accaacagac	aaccggcact	ctttccttgc	cgacccgcgt	atgcagaacc	agggccaaat	38580
gaacgagacg	ggcgcctacg	aacccatgaa	ctatcaaaac	gcgtatcatc	cgcatcaaca	38640
acaatacgaa	tctgaagacg	gggggagcag	actcagcggc	cccgtgccag	acgtgctgcg	38700
gccgggtcct	tcatccgggt	ccatagagca	gcacgaccaa	gctaacaacg	acaacaatat	38760
5gtggaataat	cgcgagtact	atggtaacag	cccatcgatat	gcaggcggat	acacgcaaga	38820
tggcaatatc	cacgagcagc	aacaacacga	tgagtacacg	agtaatgcgt	catatggcgg	38880
aatcaagga	gcaggcggag	gcagcggcgg	cgggtggcgg	ctccgagttg	caaatcgtga	38940
cagctccgac	agcaggggtg	cagatgacga	cgttgaggaga	cgtgatgcc	ttgctcagat	39000
caatthttgcg	ggcggcgctg	ctgctgcctc	cgctggagca	cctgctgctg	gtgcttcttc	39060
10ttcgcagccg	ggccatgcgc	agtagacggg	atatgcgtga	gttttttttt	aaatttcgta	39120
catagagacc	gttgatatac	caggtttcaa	attagaagag	cgaatatgca	tatcagctgt	39180
tgttcaatgt	tctagtttgg	gaaggttaac	ccccccccct	tccccttcca	agacttttca	39240
cttgthttgtg	tgtgatttaa	atctggagat	ttcaaatcta	catctcgcta	tacataggtg	39300
ttgthttgata	acgtaggggg	cagaagggta	tctcgtgata	ttagactggg	agttgcatga	39360
15atcaaggtgt	tgagcaaaaa	aagagagagc	ggtgaagggc	gggggggata	ggtggtgtgc	39420
acgtggctgg	gcgtatagcg	aaaagagcca	aaggaatgat	gacacgggac	ggcacagaag	39480
catctcggtg	gatacgaaac	atgacaacgc	cctcagtcag	caggggtgtg	ctcgaaaagt	39540
gtaggtacat	acggacgtat	gtagatatgt	tatccatttg	ccattgcatc	cttctcacat	39600
gatactaacg	tgactggacg	agataagtgt	ctctgtcgca	cgcaatatct	acgcatctca	39660
20tgacaattta	ggcgccaagt	taattgtgct	gttccccacgt	ttcagccgcg	aaccaggcgg	39720
acgggataag	gaaggagaa	cccgtcttga	attaatatth	ttccttggac	tataagatct	39780
agaatgttcg	gaagatagtc	gcccacgggt	cgacagcaag	gatgtaatag	tagtattaag	39840
cgaacgaaa	atttgacagag	aaattcaaga	aaaaggcggg	gaaggaagaa	aaaaaagatt	39900
gaaagcacat	atgtggttta	ccaagacggg	atgtatatth	acaccaatct	tgtaatgttg	39960
25gttagctgct	tgtgatgttc	ataattggct	tggcaatcta	cacacgacag	catgacaagc	40020
tgaagggcaa	aaaagcttga	tgcgtgttgg	tacacgggtg	agagattgca	aaagtgtth	40080
atgtatggaa	ggttcctcca	ggcgggttgg	tgccaaaggac	ggagattggt	ggagctgaag	40140
ttagtggth	catgtggaaa	acaggcccg	agagtcgggt	ggctthttta	ggtthttth	40200
ctthgcaaaa	atcagttagt	gaagggcgag	agcggggccag	atthcaggtc	ggtcggthcg	40260
30atgaccgact	gggacgcaa	cgggtgtcac	tgcggatacg	tatcattthga	tcgtggggac	40320
ctgggagggc	tgtgtggctt	ggagthttct	cggcaagctc	atacgcccg	tcacaggcag	40380
agtgatgtc	tgggtcgtgg	tctcgcatcg	ggggatcagt	gcaggcacag	gcgtatgtac	40440
tgtgaatcga	agcttcgtgt	tatgcgcgat	ggagtaggg	ggaggagag	ggcggtgacg	40500
atgacgatag	tgggtgatgtc	gcagctaaca	gtgatgttaa	ggcattaaag	ccggcatgctg	40560
35gacaatgctt	acgtagtagt	gtacacaaga	gcagtgtctg	caataagctt	tgtgctagat	40620
ttggagtgg	gatgcccttg	gacatggaag	cgtgtcgctg	tcaatgtgtt	acaccaaagt	40680
atagcgcgat	gcggaggaag	cgccacactc	taaaaccatg	cattgaagaa	ccgagacgtg	40740
agcgggtccc	aaggtctggc	ggaaatgaca	taggtggaac	cgcaatgtga	aggattcgac	40800
gactattcca	thttthccag	ttactgcgtc	gaathttggc	aaatgtcgac	gagctgaagc	40860
40atggcttgtg	gaggactacc	agaagcgtta	tgcgcgctgg	cagggcacac	actattgtth	40920

```

gacagcgggt ggggcccaag cangcggcgc atgaacaaac gtttcaacgt tatcgtgaag 40980
tgagccgagt cgcggcagaa ggagagcgac actccggcgc tgggtgttttg aaactggctc 41040
acaccgaagc aagcaccacg gtcaagcgat gcagttgagg caagcggcgc aagcgggacg 41100
cgatggccgc atggatgcag ccagacgggc agaagagcgg acagtcagca ggacatgttt 41160
5ttgctttgtt ttgtttcgct tggggtgtgg gtgtgaggat gagctagatt gtggctagta 41220
tggaggtaat cttgtaggtt tattgcctag aagacttggg attccgaggt tgaacggcag 41280
aagcaaatag gtcggaccgc agtgctgggt agacaagggg acaatttccc ccagattggc 41340
catgctgtcc gtcgagaagg gcgattcacg aaaagagggc gttggcccggt cgagaccggg 41400
cggccagtgt ggcggcgaag ggggcgtggg tgagcggcca gagagggcag tgggttgtga 41460
10caagcacaag cacgtgcggg tggggtgggg ggggaattgtg gcaggggcca ccgtgccgct 41520
gcgaaccaag cgcgcgcatt ggttgactgt ggtatgcatg aagagcgtat acactagcag 41580
caagagaatg cagcgccgca gggtagtaag catagggcgg cgacggcgcc tcgtggcaag 41640
tagggacgag ggctgttgag tgggtgcaggt atgctgggat gctagtagta gctcctacgt 41700
aggcgtggcc gtgtaggtgc gtggcgcaag ctgctggcgt gggtgctggc ctgctggccg 41760
15ggttgctggc ctggctgccg tcttgacaaa ggcaaatagca tagagtctgtg cccagcgccg 41820
gctttcggcc ttggtagtgc actgggcgtg tgaatagctg tcagcacgcc cgctggcggt 41880
tcgcgccatg gtggagattt tgcacgcgac atggacgacg acggcctcgg cagcgtgagg 41940
aacatgtcaa aatgaacca ggggtgcatc aaagccgttt tacctgaaca gatgagtgcg 42000
atctctgccg ggatgctga tgaagttgac tcgcttggac gacggtttgg gggcaggcta 42060
20gagccgcaca tgtcatcggc cgggcatggc gtcggggcct gcacagttcc tgcag 42115

```

<210> 60

25<211> 9

<212> PRT

<213> Artificial Sequence

<220>

30<223> Cyclization domain motif

<221> SITE

<222> 6

<223> Xaa = Asp or Glu

35

<221> SITE

<222> 9

<223> Xaa = Ser or Ala

40<221> SITE

77

<222> 2-5, 7-8

<223> Xaa = any amino acid

<400> 60

5Asp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

1

5

<210> 61

<211> 15

10<212> PRT

<213> Cochliobolus heterostrophus

<400> 61

Cys Phe Ile Ala Gly Val Val Ala Val Pro Ile Asn Ser Val Asp

15 1

5

10

15

<210> 62

<211> 15

<212> PRT

20<213> Cochliobolus heterostrophus

<400> 62

Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala Pro Ile Asp

1

5

10

15

25

<210> 63

<211> 15

<212> PRT

<213> Myxococcus xanthus

30

<400> 63

Cys Leu Tyr Ala Gly Val Val Ala Val Pro Val Tyr Pro Pro Asp

1

5

10

15

35<210> 64

<211> 14

<212> PRT

<213> Bacillus brevis

40<400> 64

78

Val Leu Lys Ala Gly Gly Tyr Val Pro Ile Asp Ile Glu Tyr

1 5 10

<210> 65

5<211> 15

<212> PRT

<213> Cochliobolus carbonum

<400> 65

10Ile Leu Lys Ala Gly Gly Val Cys Val Pro Ile Asp Pro Arg Tyr

1 5 10 15

<210> 66

<211> 15

15<212> PRT

<213> Cochliobolus carbonum

<400> 66

Val Val Gln Ala Gly Gly Val Phe Val Leu Leu Glu Pro Gly His

20 1 5 10 15

<210> 67

<211> 15

<212> PRT

25<213> Fusarium scirpi

<400> 67

Val Leu Lys Ala Gly His Ala Phe Thr Leu Ile Asp Pro Ser Asp

1 5 10 15

30

<210> 68

<211> 15

<212> PRT

<213> Fusarium scirpi

35

<400> 68

Ile Leu Lys Ala Asn Leu Ala Tyr Leu Pro Leu Asp Val Arg Ser

1 5 10 15

40<210> 69

79

<211> 15

<212> PRT

<213> Aspergillus nidulans

5<400> 69

Val Trp Lys Ser Gly Ala Ala Tyr Val Pro Ile Asp Pro Thr Tyr

1

5

10

15

<210> 70

10<211> 15

<212> PRT

<213> Aspergillus nidulans

<400> 70

15Val Trp Lys Ser Gly Gly Ala Tyr Val Pro Ile Asp Pro Gly Tyr

1

5

10

15

<210> 71

<211> 15

20<212> PRT

<213> Tolypocladium nivenm

<400> 71

Ile Leu Lys Ala His Leu Ala Tyr Leu Pro Leu Asp Ile Asn Val

25 1

5

10

15

<210> 72

<211> 15

<212> PRT

30<213> Tolypocladium nivenm

<400> 72

Ile Leu Lys Ala Gly His Ala Tyr Leu Pro Leu Asp Val Asn Val

1

5

10

15

35

<210> 73

<211> 15

<212> PRT

<213> Artificial Sequence

40

80

<220>

<223> Consensus sequence

<221> SITE

5<222> 1, 6-8, 14-15

<223> Xaa = any amino acid

<400> 73

Xaa Leu Lys Ala Gly Xaa Xaa Xaa Val Pro Ile Asp Pro Xaa Xaa
10 1 5 10 15

<210> 74

<211> 19

<212> PRT

15<213> Artificial Sequence

<220>

<223> Consensus sequence

20<221> SITE

<222> 5, 8, 13-15, 18

<223> Xaa = any amino acid

<400> 74

25Phe Thr Ser Gly Xaa Thr Gly Xaa Pro Lys Gly Val Xaa Xaa Xaa His
1 5 10 15
Arg Xaa Ile

30<210> 75

<211> 19

<212> PRT

<213> Cochliobolus heterostrophus

35<400> 75

Phe Ser Arg Ala Pro Thr Gly Asp Leu Arg Gly Val Val Leu Ser His
1 5 10 15
Arg Thr Ile

40

81

<210> 76

<211> 18

<212> PRT

<213> Cochliobolus heterostrophus

5

<400> 76

Trp Thr Tyr Trp Thr Pro Asp Gln Arg Ala Val Gln Leu Gly His Ser

1

5

10

15

Gln Ile

10

<210> 77

<211> 19

<212> PRT

15<213> Myxococcus xanthus

<400> 77

Tyr Thr Ser Gly Ser Thr Ala Asp Pro Lys Gly Val Val Leu Thr His

1

5

10

15

20Arg Asn Leu

<210> 78

<211> 19

25<212> PRT

<213> Bacillus brevis

<400> 78

Tyr Thr Ser Gly Thr Thr Gly Asn Pro Lys Gly Thr Met Leu Glu His

30 1

5

10

15

Lys Gly Ile

<210> 79

35<211> 19

<212> PRT

<213> Cochliobolus carbonum

<400> 79

40Phe Thr Ser Gly Ser Thr Gly Val Pro Lys Cys Ile Val Val Thr His

82

1	5	10	15
Ser	Gln	Ile	

5<210> 80

<211> 18

<212> PRT

<213> Cochliobolus carbonum

10<400> 80

Phe Thr Ser Gly Thr Gly Val Pro Lys Gly Ala Val Ala Thr His Gln

1	5	10	15
---	---	----	----

Ala Tyr

15

<210> 81

<211> 19

<212> PRT

<213> Fusarium scirpi

20

<400> 81

Phe Thr Ser Gly Ser Thr Gly Ile Pro Lys Gly Ile Met Ile Glu His

1	5	10	15
---	---	----	----

Arg Ser Phe

25

<210> 82

<211> 19

<212> PRT

30<213> Fusarium scirpi

<400> 82

Phe Thr Ser Gly Ser Thr Gly Lys Pro Lys Gly Val Met Ile Glu His

1	5	10	15
---	---	----	----

35Arg Ala Ile

<210> 83

<211> 19

40<212> PRT

83

<213> Aspergillus nidulans

<400> 83

Tyr Thr Ser Gly Thr Thr Gly Phe Pro Lys Gly Ile Phe Lys Gln His
5 1 5 10 15
Thr Asn Val

<210> 84

10<211> 19

<212> PRT

<213> Aspergillus nidulans

<400> 84

15Tyr Thr Ser Gly Thr Thr Gly Arg Pro Lys Gly Val Thr Val Glu His
1 5 10 15
His Gly Val

20<210> 85

<211> 19

<212> PRT

<213> Tolypocladium nivenm

25<400> 85

Phe Thr Ser Gly Ser Thr Gly Lys Pro Lys Gly Val Met Ile Glu His
1 5 10 15
Arg Gly Ile

30

<210> 86

<211> 19

<212> PRT

<213> Tolypocladium nivenm

35

<400> 86

Phe Thr Ser Gly Ser Thr Gly Lys Pro Lys Gly Val Met Ile Glu His
1 5 10 15
Arg Gly Val

40

84

<210> 87
<211> 14
<212> PRT
<213> Artificial Sequence
5
<220>
<223> Consensus sequence

<221> SITE
10<222> 4, 6, 8
<223> Xaa = any amino acid

<400> 87
Gly Glu Leu Xaa Val Xaa Gly Xaa Gly Leu Ala Arg Gly Tyr
15 1 5 10

<210> 88
<211> 14
<212> PRT
20<213> Cochliobolus heterostrophus

<400> 88
Gly Glu Ile Trp Val Asp Ser Pro Ser Leu Ser Gly Gly Phe
1 5 10
25
<210> 89
<211> 14
<212> PRT
<213> Cochliobolus heterostrophus
30
<400> 89
Gly Glu Ile Trp Val Gln Ser Glu Ala Asn Ala Tyr Ser Phe
1 5 10

35<210> 90
<211> 14
<212> PRT
<213> Myxococcus xanthus

40<400> 90

85

Gly Glu Ile Trp Val Arg Gly Pro Ser Val Ala Gln Gly Tyr

1 5 10

<210> 91

5<211> 14

<212> PRT

<213> Bacillus brevis

<400> 91

10Gly Glu Leu Cys Ile Gly Gly Glu Gly Leu Ala Arg Gly Tyr

1 5 10

<210> 92

<211> 14

15<212> PRT

<213> Cochliobolus carbonum

<400> 92

Gly Glu Leu Leu Ile Glu Ser Gly His Leu Ala Asp Lys Tyr

20 1 5 10

<210> 93

<211> 14

<212> PRT

25<213> Cochliobolus carbonum

<400> 93

Gly Glu Leu Ile Ile Glu Gly Ser Ile Leu Cys Arg Gly Tyr

1 5 10

30

<210> 94

<211> 14

<212> PRT

<213> Fusarium scirpi

35

<400> 94

Gly Glu Leu Val Ile Glu Ser Ala Gly Ile Ala Arg Asp Tyr

1 5 10

40<210> 95

86

<211> 14

<212> PRT

<213> Fusarium scirpi

5<400> 95

Gly Glu Leu Val Val Thr Gly Asp Gly Val Gly Arg Gly Tyr

1

5

10

<210> 96

10<211> 14

<212> PRT

<213> Aspergillus nidulans

<400> 96

15Gly Glu Leu His Ile Gly Gly Leu Gly Ile Ser Lys Gly Tyr

1

5

10

<210> 97

<211> 14

20<212> PRT

<213> Aspergillus nidulans

<400> 97

Gly Glu Leu Tyr Leu Gly Gly Glu Gly Val Val Arg Gly Tyr

25 1

5

10

<210> 98

<211> 14

<212> PRT

30<213> Tolypocladium nivenm

<400> 98

Gly Glu Leu Val Val Ser Gly Asp Gly Leu Ala Arg Gly Tyr

1

5

10

35

<210> 99

<211> 14

<212> PRT

<213> Tolypocladium nivenm

40

87

<400> 99

Gly Glu Leu Val Val Thr Gly Asp Gly Leu Ala Arg Gly Tyr

1

5

10

5<210> 100

<211> 8

<212> PRT

<213> Artificial Sequence

10<220>

<223> Consensus sequence

<221> SITE

<222> 7

15<223> Xaa = any amino acid

<400> 100

Tyr Arg Thr Gly Asp Leu Xaa Arg

1

5

20

<210> 101

<211> 9

<212> PRT

<213> Cochliobolus heterostrophus

25

<400> 101

Phe Leu Arg Thr Gly Leu Leu Gly Phe

1

5

30<210> 102

<211> 9

<212> PRT

<213> Cochliobolus heterostrophus

35<400> 102

Tyr Val Arg Thr Gly Asp Leu Gly Phe

1

5

<210> 103

40<211> 9

88

<212> PRT

<213> Myxococcus xanthus

<400> 103

5Trp Leu Arg Thr Gly Asp Leu Gly Phe

1 5

<210> 104

<211> 8

10<212> PRT

<213> Bacillus brevis

<400> 104

Tyr Lys Thr Gly Asp Gln Ala Arg

15 1 5

<210> 105

<211> 8

<212> PRT

20<213> Cochliobolus carbonum

<400> 105

Tyr Arg Thr Gly Asp Leu Val Arg

1 5

25

<210> 106

<211> 8

<212> PRT

<213> Cochliobolus carbonum

30

<400> 106

Tyr Lys Thr Gly Asp Leu Val Arg

1 5

35<210> 107

<211> 8

<212> PRT

<213> Fusarium scirpi

40<400> 107

89

Tyr Arg Thr Gly Asp Leu Ala Cys
1 5

<210> 108

5<211> 8

<212> PRT

<213> *Fusarium scirpi*

<400> 108

10Tyr Arg Thr Gly Asp Arg Met Arg
1 5

<210> 109

<211> 8

15<212> PRT

<213> *Aspergillus nidulans*

<400> 109

Tyr Lys Thr Gly Asp Leu Ala Arg
20 1 5

<210> 110

<211> 8

<212> PRT

25<213> *Aspergillus nidulans*

<400> 110

Tyr Lys Thr Gly Asp Leu Val Arg
1 5
30

<210> 111

<211> 8

<212> PRT

<213> *Tolypocladium nivenm*

35

<400> 111

Tyr Arg Thr Gly Asp Arg Ala Arg
1 5

40<210> 112

90

<211> 8

<212> PRT

<213> Tolypocladium nivenm

5<400> 112

Tyr Arg Thr Gly Asp Arg Ala Arg

1

5

<210> 113

10<211> 21

<212> PRT

<213> Artificial Sequence

<220>

15<223> Consensus sequence

<221> SITE

<222> 4, 6, 13

<223> Xaa = any amino acid

20

<400> 113

Leu Gly Arg Xaa Asp Xaa Gln Val Lys Ile Arg Gly Xaa Arg Ile Glu

1

5

10

15

Leu Gly Glu Val Glu

25

20

<210> 114

<211> 18

<212> PRT

30<213> Cochliobolus heterostrophus

<400> 114

Leu Gly Leu Tyr Glu Asp Arg Ile Arg Gln Arg Val Glu Asn Gly Gln

1

5

10

15

35Leu Glu

<210> 115

<211> 21

40<212> PRT

91

<213> *Bacillus brevis*

<400> 115

Leu Gly Arg Ile Asp Asn Gln Val Lys Ile Arg Gly His Arg Val Glu
5 1 5 10 15
Leu Glu Glu Val Glu
20

<210> 116

10<211> 21

<212> PRT

<213> *Cochliobolus carbonum*

<400> 116

15Leu Gly Arg Lys Asp Thr Gln Val Lys Met Asn Gly Gln Arg Phe Glu
1 5 10 15
Leu Gly Glu Val Glu
20

20<210> 117

<211> 21

<212> PRT

<213> *Cochliobolus carbonum*

25<400> 117

Val Gly Arg Ser Asp Thr Gln Ile Lys Leu Ala Gly Gln Arg Val Glu
1 5 10 15
Leu Gly Asp Val Glu
20

30

<210> 118

<211> 21

<212> PRT

<213> *Fusarium scirpi*

35

<400> 118

Leu Gly Arg Met Asp Ser Gln Val Lys Ile Arg Gly Gln Arg Val Glu
1 5 10 15
Leu Gly Ala Val Glu
40 20

92

<210> 119

<211> 21

<212> PRT

<213> *Fusarium scirpi*

5

<400> 119

Phe Gly Arg Met Asp Asn Gln Phe Lys Ile Arg Gly Asn Arg Ile Glu

1

5

10

15

Ala Gly Glu Val Glu

10

20

<210> 120

<211> 21

<212> PRT

15<213> *Aspergillus nidulans*

<400> 120

Leu Gly Arg Ala Asp Phe Gln Ile Lys Leu Arg Gly Ile Arg Ile Glu

1

5

10

15

20Pro Gly Glu Ile Glu

20

<210> 121

<211> 21

25<212> PRT

<213> *Aspergillus nidulans*

<400> 121

Leu Gly Arg Asn Asp Phe Gln Val Lys Ile Arg Gly Leu Arg Ile Glu

30 1

5

10

15

Leu Gly Glu Ile Glu

20

<210> 122

35<211> 21

<212> PRT

<213> *Tolypocladium nivenm*

<400> 122

40Phe Gly Arg Met Asp Gln Gln Val Lys Ile Arg Gly His Arg Ile Glu

93

1	5	10	15		
Pro	Ala	Glu	Val	Glu	
					20

5<210> 123

<211> 21

<212> PRT

<213> Tolypocladium nivenm

10<400> 123

Phe	Gly	Arg	Met	Asp	His	Gln	Val	Lys	Val	Arg	Gly	His	Arg	Ile	Glu
1					5					10					15
Leu	Ala	Glu	Val	Glu											
															20

15

<210> 124

<211> 21

<212> PRT

<213> Cochliobolus heterostrophus

20

<400> 124

Leu	Gly	Ser	Ile	Gly	Asp	Thr	Phe	Glu	Val	Asn	Gly	Leu	Asn	His	Phe
1					5						10				15
Ser	Met	Asp	Ile	Glu											
															20
															25

<210> 125

<211> 21

<212> PRT

30<213> Myxococcus xanthus

<400> 125

Ser	Gly	Arg	Arg	Lys	Asp	Leu	Leu	Val	Ile	Arg	Gly	Arg	Asn	Tyr	Tyr
1				5						10					15
35Pro	Gln	Asp	Leu	Glu											
															20

<210> 126

<211> 13

40<212> PRT

94

<213> Artificial Sequence

<220>

<223> Consensus sequence

5

<221> SITE

<222> 3-4, 10, 12-13

<223> Xaa = any amino acid

10<400> 126

Phe Phe Xaa Xaa Gly Gly Asp Ser Leu Xaa Ala Xaa Xaa

1

5

10

<210> 127

15<211> 13

<212> PRT

<213> Cochliobolus heterostrophus

<400> 127

20Leu Asp Ile Pro Phe Leu Asp Ser Leu Ser Glu Arg Cys

1

5

10

<210> 128

<211> 13

25<212> PRT

<213> Cochliobolus heterostrophus

<400> 128

Arg Asp Pro Asn Gly Gln Asp Ser Gln Met Ile Thr Glu

30 1

5

10

<210> 129

<211> 13

<212> PRT

35<213> Myxococcus xanthus

<400> 129

Leu Pro Asp Leu Gly Leu Asp Ser Leu Ala Leu Val Glu

1

5

10

40

95

<210> 130

<211> 13

<212> PRT

<213> *Bacillus brevis*

5

<400> 130

Phe Tyr Ala Leu Gly Gly Asp Ser Ile Lys Ala Ile Gln

1

5

10

10<210> 131

<211> 13

<212> PRT

<213> *Cochliobolus carbonum*

15<400> 131

Phe Ile His Ala Gly Gly Asp Ser Ile Thr Ala Met Gln

1

5

10

<210> 132

20<211> 13

<212> PRT

<213> *Cochliobolus carbonum*

<400> 132

25Phe Phe Ser Ser Gly Gly Asn Ser Met Ala Ala Ile Ala

1

5

10

<210> 133

<211> 13

30<212> PRT

<213> *Fusarium scirpi*

<400> 133

Phe Phe Glu Met Gly Gly Asn Ser Ile Ile Ala Ile Lys

35 1

5

10

<210> 134

<211> 13

<212> PRT

40<213> *Fusarium scirpi*

96

<400> 134

Phe Phe Gln Leu Gly Gly His Ser Leu Leu Ala Thr Lys

1 5 10

5<210> 135

<211> 13

<212> PRT

<213> *Aspergillus nidulans*

10<400> 135

Phe Phe Arg Leu Gly Gly His Ser Ile Thr Cys Ile Gln

1 5 10

<210> 136

15<211> 13

<212> PRT

<213> *Aspergillus nidulans*

<400> 136

20Phe Phe Ser Leu Gly Gly Asp Ser Leu Lys Ser Thr Lys

1 5 10

<210> 137

<211> 13

25<212> PRT

<213> *Tolypocladium nivenm*

<400> 137

Phe Phe Asp Leu Gly Gly His Ser Leu Thr Ala Met Lys

30 1 5 10

<210> 138

<211> 13

<212> PRT

35<213> *Tolypocladium nivenm*

<400> 138

Phe Phe Asn Val Gly Gly His Ser Leu Leu Ala Thr Lys

1 5 10

40

97

<210> 139
<211> 16
<212> PRT
<213> Artificial Sequence
5
<220>
<223> Consensus sequence

<221> SITE
10<222> 1-4, 6, 8, 10-12, 16
<223> Xaa = any amino acid

<400> 139
Xaa Xaa Xaa Xaa Gly Xaa Ser Xaa Gly Xaa Xaa Xaa Ala Phe Glu Xaa
15 1 5 10 15

<210> 140
<211> 16
<212> PRT
20<213> Cochliobolus heterostrophus

<400> 140
Val Leu Arg Pro Gly Pro Ser Ser Gly Ser Glu Gln His Asp Gln Ala
1 5 10 15
25
<210> 141
<211> 16
<212> PRT
<213> Aspergillus nidulans
30
<400> 141
Tyr His Phe Ile Gly Trp Ser Phe Gly Gly Thr Ile Ala Met Glu Ile
1 5 10 15

35<210> 142
<211> 16
<212> PRT
<213> Bacillus brevis

40<400> 142

98

Tyr Val Leu Ile Gly Tyr Ser Ser Gly Gly Asn Leu Ala Phe Glu Val
1 5 10 15

<210> 143

5<211> 16

<212> PRT

<213> Bacillus brevis

<400> 143

10Phe Ala Phe Leu Gly His Ser Met Gly Ala Leu Ile Ser Phe Glu Leu
1 5 10 15

<210> 144

<211> 16

15<212> PRT

<213> Myxococcus xanthus

<400> 144

Leu Thr Leu Phe Gly Tyr Ser Ala Gly Cys Ser Leu Ala Phe Glu Ala
20 1 5 10 15

<210> 145

<211> 16

<212> PRT

25<213> Brevibacillus brevis

<400> 145

Tyr Thr Leu Met Gly Tyr Ser Ser Gly Gly Asn Leu Ala Phe Glu Val
1 5 10 15
30

<210> 146

<211> 16

<212> PRT

<213> Brevibacillus brevis

35

<400> 146

Phe Ala Phe Phe Gly His Ser Met Gly Gly Leu Val Ala Phe Glu Leu
1 5 10 15

40<210> 147

99

<211> 5

<212> PRT

<213> Artificial Sequence

5<220>

<223> Consensus sequence

<221> SITE

<222> 2, 4

10<223> Xaa = any amnio acid

<400> 147

Gly Xaa Ser Xaa Gly

1

5

15

<210> 148

<211> 19

<212> DNA

<213> Artificial Sequence

20

<220>

<223> Primer

<400> 148

25gctagcatgg ccctcacac

19

<210> 149

<211> 18

<212> DNA

30<213> Artificial Sequence

<220>

<223> Primer

35<400> 149

acgatcaggg ttggagaa

18

<210> 150

<211> 17

40<212> DNA

100

<213> Artificial Sequence

<220>

<223> Primer

5

<400> 150

agcaaagcgc attcctc 17

<210> 151

10<211> 24

<212> DNA

<213> Artificial Sequence

<220>

15<223> Primer

<400> 151

gtctctatct agctacggca ttgt 24

20<210> 152

<211> 20

<212> DNA

<213> Artificial Sequence

25<220>

<223> Primer

<400> 152

gacgggccgc tagtatccat 20

30

<210> 153

<211> 23

<212> DNA

<213> Artificial Sequence

35

<220>

<223> Primer

<400> 153

40acgtctcaag tcaatgccca ata 23

101

<210> 154
<211> 24
<212> DNA
<213> Artificial Sequence
5
<220>
<223> Primer

<400> 154
10caaaactcgga tctcttctct acag 24

<210> 155
<211> 18
<212> DNA
15<213> Artificial Sequence

<220>
<223> Primer

20<400> 155
cacgatggcg gcttacag 18

<210> 156
<211> 18
25<212> DNA
<213> Artificial Sequence

<220>
<223> Primer
30
<400> 156
acacaaggcc gatgaagc 18

<210> 157
35<211> 19
<212> DNA
<213> Artificial Sequence

<220>
40<223> Primer

102

<400> 157
cgtcgacgta tccatcctt 19

<210> 158
5<211> 21
<212> DNA
<213> Artificial Sequence

<220>
10<223> Primer

<400> 158
ttccagcgcg taagtaagtc a 21

15<210> 159
<211> 23
<212> DNA
<213> Artificial Sequence

20<220>
<223> Primer

<400> 159
actcggaccg gacggaataa caa 23
25

<210> 160
<211> 19
<212> DNA
<213> Artificial Sequence

30
<220>
<223> Primer

<400> 160
35gccatgtgca gtgaagagg 19

<210> 161
<211> 17
<212> DNA
40<213> Artificial Sequence

103

<220>

<223> Primer

<400> 161

5catggcgact ccctggt

17

<210> 162

<211> 22

<212> DNA

10<213> Artificial Sequence

<220>

<223> Primer

15<400> 162

gtcatgtcta cccttcctct ca

22

<210> 163

<211> 22

20<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

25

<400> 163

acatatagtt tggacgctct gc

22

<210> 164

30<211> 24

<212> DNA

<213> Artificial Sequence

<220>

35<223> Primer

<400> 164

tatcgcccta cctacaacgc acta

24

40<210> 165

104

<211> 19

<212> DNA

<213> Artificial Sequence

5<220>

<223> Primer

<400> 165

ggacgagggc tttagtgg

19

10

<210> 166

<211> 20

<212> DNA

<213> Artificial Sequence

15

<220>

<223> Primer

<400> 166

20gcctagcaaa tggagtaaag

20

<210> 167

<211> 19

<212> DNA

25<213> Artificial Sequence

<220>

<223> Primer

30<400> 167

ggcctacccg cttctatca

19

<210> 168

<211> 20

35<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

40

105

<400> 168
agccctctgc agacgaatcc 20

<210> 169
5<211> 19
<212> DNA
<213> Artificial Sequence

<220>
10<223> Primer

<400> 169
atcaggcgag aaggtgttg 19

15<210> 170
<211> 18
<212> DNA
<213> Artificial Sequence

20<220>
<223> Primer

<400> 170
tggcggttgct tcctacag 18

25
<210> 171
<211> 19
<212> DNA
<213> Artificial Sequence

30
<220>
<223> Primer

<400> 171
35tgggcagcaa atggcacag 19

<210> 172
<211> 18
<212> DNA
40<213> Artificial Sequence

106

<220>

<223> Primer

<400> 172

5gcgctgcaca acccatca

18

<210> 173

<211> 22

<212> DNA

10<213> Artificial Sequence

<220>

<223> Primer

15<400> 173

ctgccaagga atttcatcaa gt

22

<210> 174

<211> 21

20<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

25

<400> 174

tgtgttgacc tccactagct c

21

<210> 175

30<211> 21

<212> DNA

<213> Artificial Sequence

<220>

35<223> Primer

<400> 175

cgctgacgtt tgaccatctg a

21

40<210> 176

107

<211> 21

<212> DNA

<213> Artificial Sequence

5<220>

<223> Primer

<400> 176

ctctcaaccc acaacctaac c

21

10

<210> 177

<211> 22

<212> DNA

<213> Artificial Sequence

15

<220>

<223> Primer

<400> 177

20ttctttcaaag tactcgtgtt cc

22

<210> 178

<211> 21

<212> DNA

25<213> Artificial Sequence

<220>

<223> Primer

30<400> 178

gttgcgtagt ggccgatgaa t

21

<210> 179

<211> 18

35<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

40

108

<400> 179
atgcttcgcg ggttatgg 18

<210> 180
5<211> 18
<212> DNA
<213> Artificial Sequence

<220>
10<223> Primer

<400> 180
gcgaagatgt gcgtggtg 18

15<210> 181
<211> 21
<212> DNA
<213> Artificial Sequence

20<220>
<223> Primer

<400> 181
agacccagct gttgccatt g 21

25
<210> 182
<211> 21
<212> DNA
<213> Artificial Sequence

30
<220>
<223> Primer

<400> 182
35tttgggtccg aagtagagat t 21

<210> 183
<211> 19
<212> DNA
40<213> Artificial Sequence

109

<220>

<223> Primer

<400> 183

5ggcaagaatc gaccctacc

19

<210> 184

<211> 711

<212> PRT

10<213> *Saccharomyces cerevisiae*

<400> 184

```

Met Tyr Trp Val Leu Leu Cys Gly Ser Ile Leu Leu Cys Cys Leu Ser
 1             5             10             15
15Gly Ala Ser Ala Ser Pro Ala Lys Thr Lys Met Tyr Gly Lys Leu Pro
      20             25             30
Leu Val Leu Thr Asp Ala Cys Met Gly Val Leu Gly Glu Val Thr Trp
      35             40             45
Glu Tyr Ser Ser Asp Asp Leu Tyr Ser Ser Pro Ala Cys Thr Tyr Glu
20    50             55             60
Pro Ala Leu Gln Ser Met Leu Tyr Cys Ile Tyr Glu Ser Leu Asn Glu
65             70             75             80
Lys Gly Tyr Ser Asn Arg Thr Phe Glu Lys Thr Phe Ala Ala Ile Lys
      85             90             95
25Glu Asp Cys Ala Tyr Tyr Thr Asp Asn Leu Gln Asn Met Thr Asn Ala
      100             105             110
Asp Phe Tyr Asn Met Leu Asn Asn Gly Thr Thr Tyr Ile Ile Gln Tyr
      115             120             125
Ser Glu Gly Ser Ala Asn Leu Thr Tyr Pro Ile Glu Met Asp Ala Gln
30    130             135             140
Val Arg Glu Asn Tyr Tyr Tyr Ser Tyr His Gly Phe Tyr Ala Asn Tyr
145             150             155             160
Asp Ile Gly His Thr Tyr Gly Gly Ile Ile Cys Ala Tyr Phe Val Gly
      165             170             175
35Val Met Ile Leu Ala Ser Ile Leu His Tyr Leu Ser Tyr Thr Pro Phe
      180             185             190
Lys Thr Ala Leu Phe Lys Gln Arg Leu Val Arg Tyr Val Arg Arg Tyr
      195             200             205
Leu Thr Ile Pro Thr Ile Trp Gly Lys His Ala Ser Ser Phe Ser Tyr
40    210             215             220

```

110

Leu Lys Ile Phe Thr Gly Phe Leu Pro Thr Arg Ser Glu Gly Val Ile
 225 230 235 240
 Ile Leu Gly Tyr Leu Val Leu His Thr Val Phe Leu Ala Tyr Gly Tyr
 245 250 255
 5Gln Tyr Asp Pro Tyr Asn Leu Ile Phe Asp Ser Arg Arg Glu Gln Ile
 260 265 270
 Ala Arg Tyr Val Ala Asp Arg Ser Gly Val Leu Ala Phe Ala His Phe
 275 280 285
 Pro Leu Ile Ala Leu Phe Ala Gly Arg Asn Asn Phe Leu Glu Phe Ile
 10 290 295 300
 Ser Gly Val Lys Tyr Thr Ser Phe Ile Met Phe His Lys Trp Leu Gly
 305 310 315 320
 Arg Met Met Phe Leu Asp Ala Val Ile His Gly Ala Ala Tyr Thr Ser
 325 330 335
 15Tyr Ser Val Phe Tyr Lys Asp Trp Ala Ala Ser Lys Glu Glu Thr Tyr
 340 345 350
 Trp Gln Phe Gly Val Ala Ala Leu Cys Ile Val Gly Val Met Val Phe
 355 360 365
 Phe Ser Leu Ala Met Phe Arg Lys Phe Phe Tyr Glu Ala Phe Leu Phe
 20 370 375 380
 Leu His Ile Val Leu Gly Ala Leu Phe Phe Tyr Thr Cys Trp Glu His
 385 390 395 400
 Val Val Glu Leu Ser Gly Ile Glu Trp Ile Tyr Ala Ala Ile Ala Ile
 405 410 415
 25Trp Thr Ile Asp Arg Leu Ile Arg Ile Val Arg Val Ser Tyr Phe Gly
 420 425 430
 Phe Pro Lys Ala Ser Leu Gln Leu Val Gly Asp Asp Ile Ile Arg Val
 435 440 445
 Thr Val Lys Arg Pro Val Arg Leu Trp Lys Ala Lys Pro Gly Gln Tyr
 30 450 455 460
 Val Phe Val Ser Phe Leu His His Leu Tyr Phe Trp Gln Ser His Pro
 465 470 475 480
 Phe Thr Val Leu Asp Ser Ile Ile Lys Asp Gly Glu Leu Thr Ile Ile
 485 490 495
 35Leu Lys Glu Lys Lys Gly Val Thr Lys Leu Val Lys Lys Tyr Val Cys
 500 505 510
 Cys Asn Gly Gly Lys Ala Ser Met Arg Leu Ala Ile Glu Gly Pro Tyr
 515 520 525
 Gly Ser Ser Ser Pro Val Asn Asn Tyr Asp Asn Val Leu Leu Leu Thr
 40 530 535 540

111

Gly Gly Thr Gly Leu Pro Gly Pro Ile Ala His Ala Ile Lys Leu Gly
 545 550 555 560
 Lys Thr Ser Ala Ala Thr Gly Lys Gln Phe Ile Lys Leu Val Ile Ala
 565 570 575
 5Val Arg Gly Phe Asn Val Leu Glu Ala Tyr Lys Pro Glu Leu Met Cys
 580 585 590
 Leu Glu Asp Leu Asn Val Gln Leu His Ile Tyr Asn Thr Met Glu Val
 595 600 605
 Pro Ala Leu Thr Pro Asn Asp Ser Leu Glu Ile Ser Gln Gln Asp Glu
 10 610 615 620
 Lys Ala Asp Gly Lys Gly Val Val Met Ala Thr Thr Leu Glu Gln Ser
 625 630 635 640
 Pro Asn Pro Val Glu Phe Asp Gly Thr Val Phe His His Gly Arg Pro
 645 650 655
 15Asn Val Glu Lys Leu Leu His Glu Val Gly Asp Leu Asn Gly Ser Leu
 660 665 670
 Ala Val Val Cys Cys Gly Pro Pro Val Phe Val Asp Glu Val Arg Asp
 675 680 685
 Gln Thr Ala Asn Leu Val Leu Glu Lys Pro Ala Lys Ala Ile Glu Tyr
 20 690 695 700
 Phe Glu Glu Tyr Gln Ser Trp
 705 710

<210> 185

25<211> 1774

<212> PRT

<213> Cochliobolus heterostrophus

<400> 185

30Met Met Gly Asn Tyr Ala Phe Asn Pro Asp Asn Gln Gln Ser Tyr Asp
 1 5 10 15
 Gly Gln Phe Gly Ser Pro Gly Glu Ala Ser Arg Arg Ser Thr Met Leu
 20 25 30
 Glu Val Asn Gln Gly Tyr Phe Ser Asp Phe Thr Gly Gln Gln Met Gln
 35 35 40 45
 Asp Asn Arg Asp Ser Tyr Gly Gly Pro Asn Arg Tyr Ser Ser Gly Asp
 50 55 60
 Ala Phe Ser Pro Thr Ala Ala Ile Pro Pro Pro Met Met Asn Pro Asn
 65 70 75 80
 40Asp Leu Pro Leu Gly Ala Ala Glu Thr Met Met Pro Leu Glu Pro Arg

112

				85					90				95			
	Asp	Leu	Pro	Phe	Asp	Val	Tyr	Asp	Pro	His	Asn	Pro	Asn	Val	Lys	Met
				100					105				110			
	Ser	Lys	Phe	Asp	Asn	Ile	Gly	Ala	Val	Leu	Arg	His	Arg	Ser	Arg	Thr
5			115					120					125			
	Gln	Pro	Arg	Thr	Thr	Ala	Phe	Trp	Val	Leu	Asp	Ala	Lys	Gly	Lys	Glu
			130					135					140			
	Thr	Ala	Ser	Ile	Thr	Trp	Glu	Lys	Val	Ala	Ser	Arg	Ala	Glu	Lys	Val
	145					150					155				160	
10	Ala	Lys	Val	Ile	Arg	Asp	Lys	Ser	Asn	Leu	Tyr	Arg	Gly	Asp	Arg	Val
				165						170				175		
	Ala	Leu	Val	Tyr	Arg	Asp	Thr	Glu	Ile	Ile	Asp	Phe	Val	Val	Ala	Leu
				180					185					190		
	Met	Gly	Cys	Phe	Ile	Ala	Gly	Val	Val	Ala	Val	Pro	Ile	Asn	Ser	Val
15			195					200					205			
	Asp	Asp	Tyr	Gln	Lys	Leu	Ile	Leu	Leu	Leu	Thr	Thr	Thr	Gln	Ala	His
		210					215					220				
	Leu	Ala	Leu	Thr	Thr	Asp	Asn	Asn	Leu	Lys	Ala	Phe	His	Arg	Asp	Ile
		225				230				235					240	
20	Ser	Gln	Asn	Arg	Leu	Lys	Trp	Pro	Ser	Gly	Val	Glu	Trp	Trp	Lys	Thr
				245						250					255	
	Asn	Glu	Phe	Gly	Ser	His	His	Pro	Lys	Lys	His	Asp	Asp	Thr	Pro	Ala
			260						265					270		
	Leu	Gln	Val	Pro	Glu	Val	Ala	Tyr	Ile	Glu	Phe	Ser	Arg	Ala	Pro	Thr
25			275					280						285		
	Gly	Asp	Leu	Arg	Gly	Val	Val	Leu	Ser	His	Arg	Thr	Ile	Met	His	Gln
		290				295						300				
	Met	Ala	Cys	Ile	Ser	Ala	Met	Ile	Ser	Thr	Ile	Pro	Thr	Asn	Ala	Gln
		305				310					315				320	
30	Ser	Gln	Asp	Thr	Phe	Ser	Thr	Ser	Leu	Arg	Asp	Ala	Glu	Gly	Lys	Phe
				325						330					335	
	Val	Ala	Pro	Ala	Pro	Ser	Arg	Asn	Pro	Thr	Glu	Val	Ile	Leu	Thr	Tyr
			340						345					350		
	Leu	Asp	Pro	Arg	Glu	Ser	Ala	Gly	Leu	Ile	Leu	Ser	Val	Leu	Phe	Ala
35			355					360					365			
	Val	Tyr	Gly	Gly	His	Thr	Thr	Val	Trp	Leu	Glu	Thr	Ala	Thr	Met	Glu
		370				375						380				
	Thr	Pro	Gly	Leu	Tyr	Ala	His	Leu	Ile	Thr	Lys	Tyr	Lys	Ser	Asn	Ile
		385				390					395				400	
40	Leu	Leu	Ala	Asp	Tyr	Pro	Gly	Leu	Lys	Arg	Ala	Ala	Tyr	Asn	Tyr	Gln

113

				405					410					415		
	Gln	Asp	Pro	Met	Ala	Thr	Arg	Asn	Phe	Lys	Lys	Asn	Thr	Glu	Pro	Asn
				420					425					430		
	Phe	Ala	Ser	Val	Lys	Ile	Cys	Leu	Ile	Asp	Thr	Leu	Thr	Val	Asp	Cys
5			435					440					445			
	Glu	Phe	His	Glu	Ile	Leu	Gly	Asp	Arg	Tyr	Phe	Arg	Pro	Leu	Arg	Asn
			450				455					460				
	Pro	Arg	Ala	Arg	Glu	Leu	Ile	Ala	Pro	Met	Leu	Cys	Leu	Pro	Glu	His
	465					470					475				480	
10	Gly	Gly	Met	Ile	Ile	Ser	Val	Arg	Asp	Trp	Leu	Gly	Gly	Glu	Glu	Arg
				485					490					495		
	Met	Gly	Cys	Pro	Leu	Ser	Ile	Ala	Val	Glu	Glu	Ser	Asp	Asn	Asp	Glu
			500					505					510			
	Asp	Asp	Thr	Glu	Asp	Lys	Tyr	Ala	Ala	Ala	Asn	Gly	Tyr	Ser	Ser	Leu
15			515					520					525			
	Ile	Gly	Gly	Gly	Thr	Thr	Lys	Asn	Lys	Lys	Glu	Lys	Lys	Lys	Lys	Gly
			530				535					540				
	Pro	Thr	Glu	Leu	Thr	Glu	Ile	Leu	Leu	Asp	Lys	Glu	Ala	Leu	Lys	Met
	545					550				555					560	
20	Asn	Glu	Val	Ile	Val	Leu	Ala	Ile	Gly	Glu	Glu	Ala	Ser	Lys	Arg	Ala
				565					570					575		
	Asn	Glu	Pro	Gly	Thr	Met	Arg	Val	Gly	Ala	Phe	Gly	Tyr	Pro	Ile	Pro
			580					585					590			
	Asp	Ala	Thr	Leu	Ala	Ile	Val	Asp	Pro	Glu	Thr	Ser	Leu	Leu	Cys	Ser
25			595					600					605			
	Pro	Tyr	Ser	Ile	Gly	Glu	Ile	Trp	Val	Asp	Ser	Pro	Ser	Leu	Ser	Gly
			610				615					620				
	Gly	Phe	Trp	Gln	Leu	Gln	Lys	His	Thr	Glu	Thr	Ile	Phe	His	Ala	Arg
	625				630					635					640	
30	Pro	Tyr	Arg	Phe	Val	Glu	Gly	Ser	Pro	Thr	Pro	Gln	Leu	Leu	Glu	Leu
				645					650				655			
	Glu	Phe	Leu	Arg	Thr	Gly	Leu	Leu	Gly	Phe	Val	Val	Glu	Gly	Lys	Ile
			660				665					670				
	Phe	Val	Leu	Gly	Leu	Tyr	Glu	Asp	Arg	Ile	Arg	Gln	Arg	Val	Glu	Trp
35			675				680					685				
	Val	Glu	Asn	Gly	Gln	Leu	Glu	Ala	Glu	His	Arg	Tyr	Phe	Phe	Val	Gln
			690				695					700				
	His	Leu	Val	Thr	Ser	Ile	Met	Lys	Ala	Val	Pro	Lys	Ile	Tyr	Asp	Cys
	705					710					715				720	
40	Ser	Ser	Phe	Asp	Ser	Tyr	Val	Asn	Gly	Glu	Tyr	Leu	Pro	Ile	Ile	Leu

					725					730					735	
	Ile	Glu	Thr	Gln	Ala	Ala	Ser	Thr	Ala	Pro	Thr	Asn	Pro	Gly	Gly	Pro
				740					745					750		
	Pro	Gln	Gln	Leu	Asp	Ile	Pro	Phe	Leu	Asp	Ser	Leu	Ser	Glu	Arg	Cys
5			755					760					765			
	Met	Glu	Val	Leu	Tyr	Gln	Glu	His	His	Leu	Arg	Val	Tyr	Cys	Val	Met
		770				775					780					
	Ile	Thr	Ala	Pro	Asn	Thr	Leu	Pro	Arg	Val	Ile	Lys	Asn	Gly	Arg	Arg
	785					790					795					800
10	Glu	Ile	Gly	Asn	Met	Leu	Cys	Arg	Arg	Glu	Phe	Asp	Asn	Gly	Ser	Leu
				805						810				815		
	Pro	Cys	Val	His	Val	Lys	Phe	Gly	Ile	Glu	Arg	Ser	Val	Gln	Asn	Ile
				820				825						830		
	Ala	Leu	Gly	Asp	Asp	Pro	Ala	Gly	Gly	Met	Trp	Ser	Phe	Glu	Ala	Ser
15			835					840					845			
	Met	Ala	Arg	Gln	Gln	Phe	Leu	Met	Leu	Gln	Asp	Lys	Gln	Tyr	Ser	Gly
		850				855						860				
	Val	Asp	His	Arg	Glu	Val	Val	Ile	Asp	Asp	Arg	Thr	Ser	Thr	Pro	Leu
	865					870					875					880
20	Asn	Gln	Phe	Ser	Asn	Ile	His	Asp	Leu	Met	Gln	Trp	Arg	Val	Ser	Arg
				885						890					895	
	Gln	Ala	Glu	Glu	Leu	Ala	Tyr	Cys	Thr	Val	Asp	Gly	Arg	Gly	Lys	Glu
				900					905					910		
	Gly	Lys	Gly	Val	Asn	Trp	Lys	Lys	Phe	Asp	Gln	Lys	Val	Ala	Gly	Val
25			915					920					925			
	Ala	Met	Tyr	Leu	Lys	Asn	Lys	Val	Lys	Val	Gln	Ala	Gly	Asp	His	Leu
		930				935					940					
	Leu	Leu	Met	Tyr	Thr	His	Ser	Glu	Glu	Phe	Val	Tyr	Ala	Val	His	Ala
	945					950					955					960
30	Cys	Phe	Val	Leu	Gly	Ala	Val	Cys	Ile	Pro	Met	Ala	Pro	Ile	Asp	Gln
				965						970					975	
	Asn	Arg	Leu	Asn	Glu	Asp	Ala	Pro	Ala	Leu	Leu	His	Ile	Leu	Ala	Asp
				980					985					990		
	Phe	Lys	Val	Lys	Ala	Ile	Leu	Val	Asn	Ala	Asp	Val	Asp	His	Leu	Met
35			995					1000					1005			
	Lys	Ile	Lys	Gln	Val	Ser	Gln	His	Ile	Lys	Gln	Ser	Ala	Ala	Ile	Leu
		1010				1015					1020					
	Lys	Ile	Ser	Val	Pro	Asn	Thr	Tyr	Ser	Thr	Thr	Lys	Pro	Pro	Lys	Gln
	1025					1030					1035					1040
40	Ser	Ser	Gly	Cys	Arg	Asp	Leu	Lys	Leu	Thr	Ile	Arg	Pro	Ala	Trp	Ile

115

1045 1050 1055

Gln Ala Gly Phe Pro Val Leu Val Trp Thr Tyr Trp Thr Pro Asp Gln

1060 1065 1070

Arg Arg Ile Ala Val Gln Leu Gly His Ser Gln Ile Met Ala Leu Cys

5 1075 1080 1085

Lys Val Gln Lys Glu Thr Cys Gln Met Thr Ser Thr Arg Pro Val Leu

1090 1095 1100

Gly Cys Val Arg Ser Thr Ile Gly Leu Gly Phe Leu His Thr Cys Leu

1105 1110 1115 1120

10Met Gly Ile Phe Leu Ala Ala Pro Thr Tyr Leu Val Ser Pro Val Asp

1125 1130 1135

Phe Ala Gln Asn Pro Asn Ile Leu Phe Gln Thr Leu Ser Arg Tyr Lys

1140 1145 1150

Ile Lys Asp Ala Tyr Ala Thr Ser Gln Met Leu Asp His Ala Ile Ala

15 1155 1160 1165

Arg Gly Ala Gly Lys Ser Met Ala Leu His Glu Leu Lys Asn Leu Met

1170 1175 1180

Ile Ala Thr Asp Gly Arg Pro Arg Val Asp Val Tyr Gln Arg Val Arg

1185 1190 1195 1200

20Val His Phe Ala Pro Ala Asn Leu Asp Pro Thr Ala Ile Asn Thr Val

1205 1210 1215

Tyr Ser His Val Leu Asn Pro Met Val Ala Ser Arg Ser Tyr Met Cys

1220 1225 1230

Ile Glu Pro Val Glu Leu His Leu Asp Val His Ala Leu Arg Arg Gly

25 1235 1240 1245

Leu Val Met Pro Val Asp Pro Asp Thr Glu Pro Asn Ala Leu Leu Val

1250 1255 1260

Gln Asp Ser Gly Met Val Pro Val Ser Thr Gln Ile Ser Ile Val Asn

1265 1270 1275 1280

30Pro Glu Thr Asn Gln Leu Cys Leu Asn Gly Glu Tyr Gly Glu Ile Trp

1285 1290 1295

Val Gln Ser Glu Ala Asn Ala Tyr Ser Phe Tyr Met Ser Lys Glu Arg

1300 1305 1310

Leu Asp Ala Glu Arg Phe Asn Gly Arg Thr Ile Asp Gly Asp Pro Asn

35 1315 1320 1325

Val Arg Tyr Val Arg Thr Gly Asp Leu Gly Phe Leu His Ser Val Thr

1330 1335 1340

Arg Pro Ile Gly Pro Asn Gly Ala Pro Val Asp Met Gln Val Leu Phe

1345 1350 1355 1360

40Val Leu Gly Ser Ile Gly Asp Thr Phe Glu Val Asn Gly Leu Asn His

116

	1365	1370	1375
	Phe Ser Met Asp Ile Glu Gln Ser Val Glu Arg Cys His Arg Asn Ile		
	1380	1385	1390
	Val Pro Gly Gly Cys Ala Val Phe Gln Ala Gly Gly Leu Val Val Val		
5	1395	1400	1405
	Val Val Glu Ile Phe Arg Arg Asn Phe Leu Ala Ser Met Val Pro Val		
	1410	1415	1420
	Ile Val Asn Ala Ile Leu Asn Glu His Gln Leu Val Ile Asp Ile Val		
	1425	1430	1435
10	Ser Phe Val Gln Lys Gly Asp Phe His Arg Ser Arg Leu Gly Glu Lys		
	1445	1450	1455
	Gln Arg Gly Lys Ile Leu Ala Gly Trp Val Thr Arg Lys Met Arg Thr		
	1460	1465	1470
	Ile Ala Gln Tyr Ser Ile Arg Asp Pro Asn Gly Gln Asp Ser Gln Met		
15	1475	1480	1485
	Met Ile Thr Glu Glu Pro Gly Pro Arg Ala Ser Met Thr Gly Ser Met		
	1490	1495	1500
	Leu Gly Arg Met Gly Gly Pro Ala Ser Ile Lys Ala Gly Ser Thr Arg		
	1505	1510	1515
20	Ala Pro Ser Leu Met Gly Met Thr Ala Thr Met Asn Asn Leu Ser Leu		
	1525	1530	1535
	Thr Gln Gln Gln Gln Gln Tyr Gln Gln Pro Gly Met Tyr Ala Gln		
	1540	1545	1550
	Gln Gln Gly Met His Pro Gln Gln Gln His Gln Phe Ser Met Ser Asn		
25	1555	1560	1565
	Thr Pro Pro Gln Gly Pro Pro Gln Gly Val Glu Leu His Asp Pro Ser		
	1570	1575	1580
	Asp Arg Thr Pro Thr Asp Asn Arg His Ser Phe Leu Ala Asp Pro Arg		
	1585	1590	1595
30	Met Gln Asn Gln Gly Gln Met Asn Glu Thr Gly Ala Tyr Glu Pro Met		
	1605	1610	1615
	Asn Tyr Gln Asn Ala Tyr His Pro His Gln Gln Gln Tyr Glu Ser Glu		
	1620	1625	1630
	Asp Gly Gly Ser Arg Leu Ser Gly Pro Val Pro Asp Val Leu Arg Pro		
35	1635	1640	1645
	Gly Pro Ser Ser Gly Ser Ile Glu Gln His Asp Gln Ala Asn Asn Asp		
	1650	1655	1660
	Asn Asn Met Trp Asn Asn Arg Glu Tyr Tyr Gly Asn Ser Pro Ser Tyr		
	1665	1670	1675
40	Ala Gly Gly Tyr Thr Gln Asp Gly Asn Ile His Glu Gln Gln Gln His		

117

1685 1690 1695
 Asp Glu Tyr Thr Ser Asn Ala Ser Tyr Gly Gly Asn Gln Gly Ala Gly
 1700 1705 1710
 Gly Gly Ser Gly Gly Gly Gly Gly Leu Arg Val Ala Asn Arg Asp Ser
 5 1715 1720 1725
 Ser Asp Ser Glu Gly Ala Asp Asp Asp Ala Trp Arg Arg Asp Ala Leu
 1730 1735 1740
 Ala Gln Ile Asn Phe Ala Gly Gly Ala Ala Ala Ala Ser Ala Gly Ala
 1745 1750 1755 1760
 10Pro Ala Ala Gly Ala Ser Ser Ser Gln Pro Gly His Ala Gln
 1765 1770

<210> 186

<211> 530

15<212> PRT

<213> Cochliobolus heterostrophus

<400> 186

Lys Lys Lys Gly Pro Thr Glu Leu Thr Glu Ile Leu Leu Asp Lys Glu
 20 1 5 10 15
 Ala Leu Lys Met Asn Glu Val Ile Val Leu Ala Ile Gly Glu Glu Ala
 20 25 30
 Ser Lys Arg Ala Asn Glu Pro Gly Thr Met Arg Val Gly Ala Phe Gly
 35 40 45
 25Tyr Pro Ile Pro Asp Ala Thr Leu Ala Ile Val Asp Pro Glu Thr Ser
 50 55 60
 Leu Leu Cys Ser Pro Tyr Ser Ile Gly Glu Ile Trp Val Asp Ser Pro
 65 70 75 80
 Ser Leu Ser Gly Gly Phe Trp Gln Leu Gln Lys His Thr Glu Thr Ile
 30 85 90 95
 Phe His Ala Arg Pro Tyr Arg Phe Val Glu Gly Ser Pro Thr Pro Gln
 100 105 110
 Leu Leu Glu Leu Glu Phe Leu Arg Thr Gly Leu Leu Gly Phe Val Val
 115 120 125
 35Glu Gly Lys Ile Phe Val Leu Gly Leu Tyr Glu Asp Arg Ile Arg Gln
 130 135 140
 Arg Val Glu Trp Val Glu Asn Gly Gln Leu Glu Ala Glu His Arg Tyr
 145 150 155 160
 Phe Phe Val Gln His Leu Val Thr Ser Ile Met Lys Ala Val Pro Lys
 40 165 170 175

118

Ile Tyr Asp Cys Ser Ser Phe Asp Ser Tyr Val Asn Gly Glu Tyr Leu			
180	185	190	
Pro Ile Ile Leu Ile Glu Thr Gln Ala Ala Ser Thr Ala Pro Thr Asn			
195	200	205	
5Pro Gly Gly Pro Pro Gln Gln Leu Asp Ile Pro Phe Leu Asp Ser Leu			
210	215	220	
Ser Glu Arg Cys Met Glu Val Leu Tyr Gln Glu His His Leu Arg Val			
225	230	235	240
Tyr Cys Val Met Ile Thr Ala Pro Asn Thr Leu Pro Arg Val Ile Lys			
10	245	250	255
Asn Gly Arg Arg Glu Ile Gly Asn Met Leu Cys Arg Arg Glu Phe Asp			
260	265	270	
Asn Gly Ser Leu Pro Cys Val His Val Lys Phe Gly Ile Glu Arg Ser			
275	280	285	
15Val Gln Asn Ile Ala Leu Gly Asp Asp Pro Ala Gly Gly Met Trp Ser			
290	295	300	
Phe Glu Ala Ser Met Ala Arg Gln Gln Phe Leu Met Leu Gln Asp Lys			
305	310	315	320
Gln Tyr Ser Gly Val Asp His Arg Glu Val Val Ile Asp Asp Arg Thr			
20	325	330	335
Ser Thr Pro Leu Asn Gln Phe Ser Asn Ile His Asp Leu Met Gln Trp			
340	345	350	
Arg Val Ser Arg Gln Ala Glu Glu Leu Ala Tyr Cys Thr Val Asp Gly			
355	360	365	
25Arg Gly Lys Glu Gly Lys Gly Val Asn Trp Lys Lys Phe Asp Gln Lys			
370	375	380	
Val Ala Gly Val Ala Met Tyr Leu Lys Asn Lys Val Lys Val Gln Ala			
385	390	395	400
Gly Asp His Leu Leu Leu Met Tyr Thr His Ser Glu Glu Phe Val Tyr			
30	405	410	415
Ala Val His Ala Cys Phe Val Leu Gly Ala Val Cys Ile Pro Met Ala			
420	425	430	
Pro Ile Asp Gln Asn Arg Leu Asn Glu Asp Ala Pro Ala Leu Leu His			
435	440	445	
35Ile Leu Ala Asp Phe Lys Val Lys Ala Ile Leu Val Asn Ala Asp Val			
450	455	460	
Asp His Leu Met Lys Ile Lys Gln Val Ser Gln His Ile Lys Gln Ser			
465	470	475	480
Ala Ala Ile Leu Lys Ile Ser Val Pro Asn Thr Tyr Ser Thr Thr Lys			
40	485	490	495

119

Pro Pro Lys Gln Ser Ser Gly Cys Arg Asp Leu Lys Leu Thr Ile Arg
 500 505 510
 Pro Ala Trp Ile Gln Ala Gly Phe Pro Val Leu Val Trp Thr Tyr Trp
 515 520 525
 5Thr Pro
 530

<210> 187
 <211> 1767
 10<212> DNA
 <213> *Cochliobolus heterostrophus*

<400> 187

atgggtctctt	caatttcgag	ggtgggttacc	gcccttgggc	ttcttttgcc	cactgtcact	60
15tcattctgtgt	acctcataaa	agtctcgccc	ccagaacacc	ggagcaaacg	acaaggatat	120
gagactgcct	gtaatcatgg	tccagaatca	agaggatgct	ggattgacga	cttcaacatc	180
gacaccgaca	tggatgttga	atggccagat	actggaaaga	cagtcaagta	tcacctgacc	240
atcaccaata	ccactggagc	tccagacggt	tttgaaaggc	cgatgttctt	gattaatggc	300
caatacccag	gaccaactat	tactgccgac	tggggagatg	ttctagagat	cacagttacc	360
20aatggccttg	aaaacaacgg	tacaggtata	cattggcacg	gtctgaggca	actcgggaca	420
aacgaacaag	atggcgtaaa	tggtatcact	gaatgcccaa	tcgcacccgg	tgactccaag	480
ctctacagat	tcaaagcaac	tcaatatggc	actacctggg	atcactcgca	ctactcgggtg	540
cagtatggtg	acggcatcgt	gggtcctctg	atcatcaaag	gaccctcaac	ggcgaactac	600
gatattgatc	ttggcgcttt	cccaatgact	gactgggtttc	acgcaaccac	cttcaccgtc	660
25aacgctgcag	ccgttcatgc	aaatggccct	ccaactgctg	acaatgtcct	tgtcaatggc	720
tccatgacct	catcttttgg	cggcaagtac	gccgaaacga	tcctaactcc	gggaaaatct	780
cacttgctgc	gtttgatgaa	cgttggtatt	aacaactacc	ttcatgtcgg	cctcgatggg	840
catcagttcc	aggtcatttc	ggctgatttc	acgcccattg	aacctttcta	cacggacagc	900
ttggtccttg	cagtcgggtca	acggtatgaa	gtcatcatca	acgcaactga	agctgtgggc	960
30aactactggc	tacgtgttgg	taccggcggt	aactgcgacg	gtcccaatgc	caatgcagca	1020
aatatcagga	gtatcttccg	atatgctggc	gctccaactg	aagacccaga	cacgactggg	1080
tcgcttccgt	cgggctgcta	cgatgaggat	gttgtaccct	atgccaaagac	gactgttcct	1140
caggagatgc	ccgaacagtt	gagcgtgggc	ttcaacccta	actggactag	tgacgtgacg	1200
caaaatcagg	gtctggtcca	atggctcgtc	aacggtaatc	ccatggcagt	tgatcttgaa	1260
35gtccctactc	tgcagtcggg	gttggatggc	aatgttacct	acggaaaaca	ccgccacgtg	1320
tttgcagtcg	acgagaaaca	ccaatggcaa	tattgggtca	tccaacaaaa	cagttctaac	1380
ccaccacttc	ctcaccocat	ccacctccac	ggccacgact	tctacgtcct	cgcacagggtc	1440
gaaaacgcag	tctggaacgg	agatattttca	accctgaaga	cggacaaccc	catccgtcgg	1500
gacacggccg	atcttcccgc	tggaggctac	ttggtccttg	ctttcgagtc	ggacaaccct	1560
40ggcgcacatggc	ttatgcactg	ccacatcccc	ttccacgttg	ctgccggtct	cgggtgtccag	1620

120

```

ttcctcgagc gcgaatccga aatcaaggcc caagatggat acgcagagat gcacaggaca 1680
tgtgctaact ggcagtcatt gcgctacaag taccatccca atggcatcctt gttccccggg 1740
gactctggtc tacgtcgtcg caactaa 1767

```

5<210> 188

<211> 588

<212> PRT

<213> *Cochliobolus heterostrophus*

10<400> 188

```

Met Val Ser Ser Ile Ser Arg Val Val Thr Ala Leu Gly Leu Leu Leu
  1              5              10              15
Pro Thr Val Thr Ser Ser Val Tyr Leu Ile Lys Val Ser Pro Pro Glu
      20              25              30
15His Arg Ser Lys Arg Gln Gly Tyr Glu Thr Ala Cys Asn His Gly Pro
      35              40              45
Glu Ser Arg Gly Cys Trp Ile Asp Asp Phe Asn Ile Asp Thr Asp Met
      50              55              60
Asp Val Glu Trp Pro Asp Thr Gly Lys Thr Val Lys Tyr His Leu Thr
2065              70              75              80
Ile Thr Asn Thr Thr Gly Ala Pro Asp Gly Phe Glu Arg Pro Met Phe
      85              90              95
Leu Ile Asn Gly Gln Tyr Pro Gly Pro Thr Ile Thr Ala Asp Trp Gly
      100             105             110
25Asp Val Leu Glu Ile Thr Val Thr Asn Gly Leu Glu Asn Asn Gly Thr
      115             120             125
Gly Ile His Trp His Gly Leu Arg Gln Leu Gly Thr Asn Glu Gln Asp
      130             135             140
Gly Val Asn Gly Ile Thr Glu Cys Pro Ile Ala Pro Gly Asp Ser Lys
30145             150             155             160
Leu Tyr Arg Phe Lys Ala Thr Gln Tyr Gly Thr Thr Trp Tyr His Ser
      165             170             175
His Tyr Ser Val Gln Tyr Gly Asp Gly Ile Val Gly Pro Leu Ile Ile
      180             185             190
35Lys Gly Pro Ser Thr Ala Asn Tyr Asp Ile Asp Leu Gly Ala Phe Pro
      195             200             205
Met Thr Asp Trp Phe His Ala Thr Thr Phe Thr Val Asn Ala Ala Ala
      210             215             220
Val His Ala Asn Gly Pro Pro Thr Ala Asp Asn Val Leu Val Asn Gly
40225             230             235             240

```


121

Ser Met Thr Ser Ser Phe Gly Gly Lys Tyr Ala Glu Thr Ile Leu Thr			
	245	250	255
Pro Gly Lys Ser His Leu Leu Arg Leu Met Asn Val Gly Ile Asn Asn			
	260	265	270
5Tyr Leu His Val Gly Leu Asp Gly His Gln Phe Gln Val Ile Ser Ala			
	275	280	285
Asp Phe Thr Pro Ile Glu Pro Phe Tyr Thr Asp Ser Leu Val Leu Ala			
	290	295	300
Val Gly Gln Arg Tyr Glu Val Ile Ile Asn Ala Thr Glu Ala Val Gly			
10305	310	315	320
Asn Tyr Trp Leu Arg Val Gly Thr Gly Gly Asn Cys Asp Gly Pro Asn			
	325	330	335
Ala Asn Ala Ala Asn Ile Arg Ser Ile Phe Arg Tyr Ala Gly Ala Pro			
	340	345	350
15Thr Glu Asp Pro Asp Thr Thr Gly Ser Leu Pro Ser Gly Cys Tyr Asp			
	355	360	365
Glu Asp Val Val Pro Tyr Ala Lys Thr Thr Val Pro Gln Glu Met Pro			
	370	375	380
Glu Gln Leu Ser Val Gly Phe Asn Pro Asn Trp Thr Ser Asp Val Thr			
20385	390	395	400
Gln Asn Gln Gly Leu Val Gln Trp Leu Val Asn Gly Asn Pro Met Ala			
	405	410	415
Val Asp Leu Glu Val Pro Thr Leu Gln Ser Val Leu Asp Gly Asn Val			
	420	425	430
25Thr Tyr Gly Asn Asn Arg His Val Phe Ala Val Asp Glu Lys His Gln			
	435	440	445
Trp Gln Tyr Trp Val Ile Gln Gln Asn Ser Ser Asn Pro Pro Leu Pro			
	450	455	460
His Pro Ile His Leu His Gly His Asp Phe Tyr Val Leu Ala Gln Val			
30465	470	475	480
Glu Asn Ala Val Trp Asn Gly Asp Ile Ser Thr Leu Lys Thr Asp Asn			
	485	490	495
Pro Ile Arg Arg Asp Thr Ala Asp Leu Pro Ala Gly Gly Tyr Leu Val			
	500	505	510
35Leu Ala Phe Glu Ser Asp Asn Pro Gly Ala Trp Leu Met His Cys His			
	515	520	525
Ile Pro Phe His Val Ala Ala Gly Leu Gly Val Gln Phe Leu Glu Arg			
	530	535	540
Glu Ser Glu Ile Lys Ala Gln Asp Gly Tyr Ala Glu Met His Arg Thr			
40545	550	555	560

122

Cys Ala Asn Trp Gln Ser Trp Arg Tyr Lys Tyr His Pro Asn Gly Ile

565

570

575

Leu Phe Pro Gly Asp Ser Gly Leu Arg Arg Arg Asn

580

585

5

<210> 189

<211> 327

<212> DNA

<213> Cochliobolus heterostrophus

10

<400> 189

atggcgcaag agaagaagga agaacaaccc cagcaagacc acatccccac ctgcgcgcag 60

aacgaagagg aggaacaaag caaaggctcc ggcggcctct tgagcgcaat cggagatcca 120

gtcgggtacgt ctctttatcc ccccttcctc ctccatctct caaccacaa cctaaccat 180

15ctcccaaggc aacgtcctca acaccgccct ccgccccgtc ggcgcgccgc tcgagaaatt 240

cgtcacaggc ccgctgggcg agggctctcg cggcaccaca cgcggcgcgc tgggcccgtt 300

gatgggccac gaggacgagc gctctga 327

<210> 190

20<211> 108

<212> PRT

<213> Cochliobolus heterostrophus

<400> 190

25Met Ala Gln Glu Lys Lys Glu Glu Gln Pro Gln Gln Asp His Ile Pro

1

5

10

15

Thr Ser Pro Gln Asn Glu Glu Glu Glu Gln Ser Lys Gly Ser Gly Gly

20

25

30

Leu Leu Ser Ala Ile Gly Asp Pro Val Gly Thr Ser Pro Tyr Pro Pro

30

35

40

45

Phe Leu Leu His Leu Ser Thr His Asn Leu Thr His Leu Pro Arg Gln

50

55

60

Arg Pro Gln His Arg Pro Pro Pro Arg Arg Arg Ala Ala Arg Glu Ile

65

70

75

80

35Arg His Arg Pro Ala Gly Arg Gly Ser Arg Arg His His Thr Arg Arg

85

90

95

Ala Gly Pro Val Asp Gly Pro Arg Gly Arg Ala Leu

100

105

40<210> 191

123

<211> 1626

<212> DNA

<213> Cochliobolus heterostrophus

5<400> 191

```

atggataccc tgcctgcttc ctgccggctg ctcacagctc ttccctcacg gctatgcgcc      60
cggcgatctc ttccctccaca attgcgggct gcgcgaactc tgccaccacg atggcggcctt    120
acaggtcaat tgcgctgcat cagtgagegc gcgccgacaa ttgaccgctc gaaatcaaag      180
cttttcaaag atgcagacga agcagttgca gatgtacagc ctggttccac cgtcctgagc      240
10gcaggattcg ggttgtgtgg tgtcgcagac actttgatcg cagcgatgaa gaagcgtggg      300
ccggagtcac tacattcggt aacagccgtc tcaaacaatg ctggcattga agacgtagga      360
ggattggcac atcttacaaa gaacggacaa gtcaagaagc tcattataag ttttctcggc      420
aacaacaagg cgcttgagaa gcagtatctg agcgggtggt ttgaaattga gctttgtccg      480
caaggtacgc ttgcagagag gatacgcgct ggtggcgagc gcatcccagc attttacaca      540
15ccactgcag taaatacact gttgcaagat ggccagattc cggccaagtt tgacaaggag      600
ggcaaggctg taggctacgg acagaagcgt gaggttagag agttcaatgg caagaagtcc      660
ctcatggaga ctgcattgac cggcgatgtc gccattatcc gtgcacacaa ggccgatgaa      720
gctggttaact gtgttttcag atacaccacc aaagcttttg gacccatcat ggccaaagcc      780
gcacgcctta caattgtcga agccgaagag attgtcccta taggcacctt tgatgccaac      840
20gaggtcgatc tccctggcat ctctgtcgac cgcacgtccc cagccaccgc cccaagaac      900
attgagatca agaagctgcg aaaaccggct gcatccaaag acgcctcatc aaagaacgaa      960
gccgccgaac gacgcgatcg tatcgctcgt cgggcagcaa aggagctcaa gcagggatac    1020
tacgtcaatc tgggcgtcgg catccccaca gccgcagcag cattcgtacc cgatggcgctc    1080
aagggtgtggc tgcaatccga aaatggcatt ctaggaatgg gccctaccc gacggaagaa    1140
25gaagtagacg cagacattgt caacgccggc aaagaaaccg taaccctcct tcccggcgcc    1200
tcgacctttg acagcgccga atcctttggc atgatccgcg gcggccacgt cgacgtatcc    1260
atccttggag ctctacaagt cagtgcctct ggcgacctgg ccaactacat ggtgcccggc    1320
aaagtcttca agggtatggg cggcgccatg gatctcgtta gcaatcccga tgctacaaaa    1380
gtcgttgtcg cgactgaaca cgttgctaaa gatggatcca gcaagattgt tcaggaatgc    1440
30cagttgccgc ttacaggagc aaagtgcgtg agcactatta ttaccgatct gtgtgtcttt    1500
gaagtaaaca ggaagagggg gactttgacg ctgacggaga cggcgccggg ggtagcggtt    1560
gaggatgtca aggcgaagac ggatgcgaag tttgaagtcg cgagtgatct caagacgatg    1620
gagtag                                           1626

```

35<210> 192

<211> 541

<212> PRT

<213> Cochliobolus heterostrophus

40<400> 192

124

Met	Asp	Thr	Leu	Pro	Ala	Ser	Cys	Arg	Leu	Leu	Thr	Ala	Leu	Pro	Ser
1				5					10					15	
Arg	Leu	Cys	Ala	Arg	Arg	Ser	Leu	Pro	Pro	Gln	Leu	Arg	Ala	Ala	Arg
			20					25					30		
5Thr	Leu	Pro	Pro	Arg	Trp	Arg	Leu	Thr	Gly	Gln	Leu	Arg	Cys	Ile	Ser
		35					40					45			
Glu	Arg	Ala	Pro	Thr	Ile	Asp	Arg	Ser	Lys	Ser	Lys	Leu	Phe	Lys	Asp
	50					55					60				
Ala	Asp	Glu	Ala	Val	Ala	Asp	Val	Gln	Pro	Gly	Ser	Thr	Val	Leu	Ser
1065					70					75				80	
Ala	Gly	Phe	Gly	Leu	Cys	Gly	Val	Ala	Asp	Thr	Leu	Ile	Ala	Ala	Met
				85					90					95	
Lys	Lys	Arg	Gly	Pro	Glu	Ser	Leu	His	Ser	Leu	Thr	Ala	Val	Ser	Asn
			100					105					110		
15Asn	Ala	Gly	Ile	Glu	Asp	Val	Gly	Gly	Leu	Ala	His	Leu	Thr	Lys	Asn
			115					120					125		
Gly	Gln	Val	Lys	Lys	Leu	Ile	Ile	Ser	Phe	Leu	Gly	Asn	Asn	Lys	Ala
		130				135					140				
Leu	Glu	Lys	Gln	Tyr	Leu	Ser	Gly	Gly	Ile	Glu	Ile	Glu	Leu	Cys	Pro
20145					150					155				160	
Gln	Gly	Thr	Leu	Ala	Glu	Arg	Ile	Arg	Ala	Gly	Gly	Ala	Gly	Ile	Pro
				165					170					175	
Ala	Phe	Tyr	Thr	Pro	Thr	Ala	Val	Asn	Thr	Leu	Leu	Gln	Asp	Gly	Gln
			180					185					190		
25Ile	Pro	Ala	Lys	Phe	Asp	Lys	Glu	Gly	Lys	Ala	Val	Gly	Tyr	Gly	Gln
		195					200					205			
Lys	Arg	Glu	Val	Arg	Glu	Phe	Asn	Gly	Lys	Lys	Phe	Leu	Met	Glu	Thr
	210					215					220				
Ala	Leu	Thr	Gly	Asp	Val	Ala	Ile	Ile	Arg	Ala	His	Lys	Ala	Asp	Glu
30225					230					235				240	
Ala	Gly	Asn	Cys	Val	Phe	Arg	Tyr	Thr	Thr	Lys	Ala	Phe	Gly	Pro	Ile
			245						250				255		
Met	Ala	Lys	Ala	Ala	Arg	Leu	Thr	Ile	Val	Glu	Ala	Glu	Glu	Ile	Val
		260						265					270		
35Pro	Ile	Gly	Thr	Phe	Asp	Ala	Asn	Glu	Val	Asp	Leu	Pro	Gly	Ile	Phe
		275					280						285		
Val	Asp	Arg	Ile	Val	Pro	Ala	Thr	Ala	Pro	Lys	Asn	Ile	Glu	Ile	Lys
	290					295					300				
Lys	Leu	Arg	Lys	Pro	Ala	Ala	Ser	Lys	Asp	Ala	Ser	Ser	Lys	Asn	Glu
40305					310					315				320	

125

Ala Ala Glu Arg Arg Asp Arg Ile Ala Arg Arg Ala Ala Lys Glu Leu
 325 330 335
 Lys Gln Gly Tyr Tyr Val Asn Leu Gly Val Gly Ile Pro Thr Ala Ala
 340 345 350
 5Ala Ala Phe Val Pro Asp Gly Val Lys Val Trp Leu Gln Ser Glu Asn
 355 360 365
 Gly Ile Leu Gly Met Gly Pro Tyr Pro Thr Glu Glu Glu Val Asp Ala
 370 375 380
 Asp Ile Val Asn Ala Gly Lys Glu Thr Val Thr Leu Leu Pro Gly Ala
 10385 390 395 400
 Ser Thr Phe Asp Ser Ala Glu Ser Phe Gly Met Ile Arg Gly Gly His
 405 410 415
 Val Asp Val Ser Ile Leu Gly Ala Leu Gln Val Ser Ala Ser Gly Asp
 420 425 430
 15Leu Ala Asn Tyr Met Val Pro Gly Lys Val Phe Lys Gly Met Gly Gly
 435 440 445
 Ala Met Asp Leu Val Ser Asn Pro Asp Ala Thr Lys Val Val Val Ala
 450 455 460
 Thr Glu His Val Ala Lys Asp Gly Ser Ser Lys Ile Val Gln Glu Cys
 20465 470 475 480
 Gln Leu Pro Leu Thr Gly Ala Lys Cys Val Ser Thr Ile Ile Thr Asp
 485 490 495
 Leu Cys Val Phe Glu Val Asn Arg Lys Arg Gly Thr Leu Thr Leu Thr
 500 505 510
 25Glu Thr Ala Pro Gly Val Ser Val Glu Asp Val Lys Ala Lys Thr Asp
 515 520 525
 Ala Lys Phe Glu Val Ala Ser Asp Leu Lys Thr Met Glu
 530 535 540

30<210> 193

<211> 1131

<212> DNA

<213> Cochliobolus heterostrophus

35<220>

<221> misc_feature

<222> (1)...(1131)

<223> n = any nucleotide

40<400> 193

126

```

atgaacataa agacatggct acccccgaaa acgtctgggg cggctggaat gaaactgaaa      60
tcaacaatct gcatgctcat cagaagacnt gctaaaccgc gttggaatcg tggcactcag      120
ccgtacaaga ggaagccttg gccaaaacaa agggacatga aatataattcc tggcaaaagc      180
gagagcgatg gtggtggtgt caactgctgg tctgacagca acggagaccc tgactacgat      240
5gtcaggaaac tgctagactg gaacggcgat tggctacctg ctccggaatc atgggtccgct      300
cgaagaggac atgaagaccg tcaccttggg gcacatgtag aacaatggat gaatggacac      360
tcacaagagt gcaccagatc cgtatactac ccactcagta ctttcagtcc cgaagatgga      420
ccttgcaaag agctggcacc tcgttactgg cttgaggcga aggttgaggg cagtaacttg      480
agagaatctt ggaagacaat ctctacttcg gacccaaagc cgctggatga tacggacatt      540
10actatccatc caccttggtg ggaattgtac gaggatgttg tctattctga ggtgattcac      600
gaggaagggtc aggggtgaaca gcatttcaag cataggagct gttacctgaa cagcctacca      660
gcgccggagg caagaatcga ccctaccgat gcagagcatc ctaccactca tctgatgctg      720
gcttcggctg cagaaaagct tcaagatcta caacaacgta gggaagctaa ggaacgtcgc      780
ttgttggcca aacggaatcg cccagtcgcg aattcgatgt ttccaatgca agccatggaa      840
15gatcgtcgcc tacgccctaa gaccaacatg tacattcgtc ctgttcagcc agcagatgtt      900
gttggcattg gaacaaggat gcaaagtttt caaactaaca atattgacag gcgatttaca      960
actactacgt tgagcatacc atttacgcaa ccgagtttga tgggcgcact gaagatcaaa     1020
tccgccagcg aatcaacact gtcaccagtg caggccttcc atacttggtc gcagtctcaa     1080
agagcaacga gtccaggacc aatcccgggt atgttaccga aaagattgta g              1131

```

20

<210> 194

<211> 376

<212> PRT

<213> Cochliobolus heterostrophus

25

<220>

<221> SITE

<222> (1)...(376)

<223> Xaa = any amino acid

30

<400> 194

Met Asn Ile Lys Thr Trp Leu Pro Pro Lys Thr Ser Gly Ala Ala Gly

1 5 10 15

Met Lys Leu Lys Ser Thr Ile Cys Met Leu Ile Arg Arg Xaa Ala Lys

35 20 25 30

Pro Arg Trp Asn Arg Gly Thr Gln Pro Tyr Lys Arg Lys Pro Trp Pro

35 40 45

Lys Gln Arg Asp Met Lys Tyr Ile Pro Gly Lys Ser Glu Ser Asp Gly

50 55 60

40Gly Gly Val Asn Cys Trp Ser Asp Ser Asn Gly Asp Pro Asp Tyr Asp

127

65		70		75		80										
Val	Arg	Lys	Leu	Leu	Asp	Trp	Asn	Gly	Asp	Trp	Leu	Pro	Ala	Pro	Glu	
		85		90		95										
Ser	Trp	Ser	Ala	Arg	Arg	Gly	His	Glu	Asp	Arg	His	Leu	Gly	Ala	His	
5		100		105		110										
Val	Glu	Gln	Trp	Met	Asn	Gly	His	Ser	Gln	Glu	Cys	Thr	Arg	Ser	Val	
		115		120		125										
Tyr	Tyr	Pro	Leu	Ser	Thr	Phe	Ser	Pro	Glu	Asp	Gly	Pro	Cys	Lys	Glu	
	130			135		140										
10	Leu	Ala	Pro	Arg	Tyr	Trp	Leu	Glu	Ala	Lys	Val	Glu	Gly	Ser	Asn	Leu
	145			150		155										
Arg	Glu	Ser	Trp	Lys	Thr	Ile	Ser	Thr	Ser	Asp	Pro	Lys	Pro	Leu	Asp	
		165		170		175										
Asp	Thr	Asp	Ile	Thr	Ile	His	Pro	Pro	Trp	Trp	Glu	Leu	Tyr	Glu	Asp	
15		180		185		190										
Val	Val	Tyr	Ser	Glu	Val	Ile	His	Glu	Glu	Gly	Gln	Gly	Glu	Gln	His	
	195			200		205										
Phe	Lys	His	Arg	Ser	Cys	Tyr	Leu	Asn	Ser	Leu	Pro	Ala	Pro	Glu	Ala	
	210			215		220										
20	Arg	Ile	Asp	Pro	Thr	Asp	Ala	Glu	His	Pro	Thr	Thr	His	Leu	Met	Leu
	225			230		235										
Ala	Ser	Ala	Ala	Glu	Lys	Leu	Gln	Asp	Leu	Gln	Gln	Arg	Arg	Glu	Ala	
		245		250		255										
Lys	Glu	Arg	Arg	Leu	Leu	Ala	Lys	Arg	Asn	Arg	Pro	Val	Ala	Asn	Ser	
25		260		265		270										
Met	Phe	Pro	Met	Gln	Ala	Met	Glu	Asp	Arg	Arg	Leu	Arg	Pro	Lys	Thr	
	275			280		285										
Asn	Met	Tyr	Ile	Arg	Pro	Val	Gln	Pro	Ala	Asp	Val	Val	Gly	Ile	Gly	
	290			295		300										
30	Thr	Arg	Met	Gln	Ser	Phe	Gln	Thr	Asn	Asn	Ile	Asp	Arg	Arg	Phe	Thr
	305			310		315										
Thr	Thr	Thr	Leu	Ser	Ile	Pro	Phe	Thr	Gln	Pro	Ser	Leu	Met	Gly	Ala	
		325		330		335										
Leu	Lys	Ile	Lys	Ser	Ala	Ser	Glu	Ser	Thr	Leu	Ser	Pro	Val	Gln	Ala	
35		340		345		350										
Phe	His	Thr	Trp	Ser	Gln	Ser	Gln	Arg	Ala	Thr	Ser	Pro	Gly	Pro	Ile	
	355			360		365										
Pro	Val	Met	Leu	Pro	Lys	Arg	Leu									
	370			375												

128

<210> 195

<211> 768

<212> DNA

<213> Cochliobolus heterostrophus

5

<400> 195

```

atggagaaca tggagatatc ccagcaaadc aaatccacga cattgtctgt tccctcgccg      60
accgcgacac atactgcctg tgtcaacggt gcacgtttgc aaatccgatg tctcaataact    120
ttcgagggtg tccgtactat cgccttccca tccacccatg atttgcgctc gtcgaagatt    180
10acctgggtcac ccctgggtcat tccgcccttg acctcatcaa cacgcacatc ttcgcccacc   240
actacacctc cccgtcgatc atcacggaca ccacgtccct gctcgaatcg cgtcctcata   300
tccgacgacg acaccgcgcg cgtttacgat ctccgcgatg agaaatggaa tgccgtgatt   360
agcaatggct ctggtggcat ggggaagaat gttcacgctg agtttggagg aacagaggac   420
gagggtgcttg tttggaccga ctttaccgcc tgtgttaaga tatggtgctt gaagacgggt   480
15cgggtagtgag agatacgaga tccgaagttt cctggtaaag atggcaaggg gtgggggttac   540
cgacctgctg acgatactgg attgaggaat ggaaggggac aagggcgtgt tctggcatta   600
ttgtgtcgtg catcagggac cgatatcttg ttgcttcttg caccgcagac gtacaagggt   660
ctgaatcgag tcgaactccc tactacagac gccgctggtc tgagatggag tcgtgacggg   720
cgctggctgg ccatctggga cgctgcgtct gcggggttaca agctttga      768

```

20

<210> 196

<211> 255

<212> PRT

<213> Cochliobolus heterostrophus

25

<400> 196

```

Met Glu Asn Met Glu Ile Ser Gln Gln Ile Lys Ser Thr Thr Leu Ser
  1             5             10             15
Val Pro Ser Pro Thr Ala Thr His Thr Ala Cys Val Asn Gly Ala Arg
30             20             25             30
Leu Gln Ile Arg Cys Leu Asn Thr Phe Glu Val Val Arg Thr Ile Ala
      35             40             45
Leu Pro Ser Thr His Asp Leu Arg Ser Ser Lys Ile Thr Trp Ser Pro
      50             55             60
35Leu Val Ile Pro Pro Leu Thr Ser Ser Thr Arg Thr Ser Ser Pro Thr
      65             70             75             80
Thr Thr Pro Pro Arg Arg Ser Ser Arg Thr Pro Arg Pro Cys Ser Asn
      85             90             95
Arg Val Leu Ile Ser Asp Asp Asp Thr Ala Arg Val Tyr Asp Leu Arg
40             100             105             110

```


129

Asp Glu Lys Trp Asn Ala Val Ile Ser Asn Gly Ser Gly Gly Met Gly
 115 120 125
 Lys Asn Val His Val Glu Phe Gly Gly Thr Glu Asp Glu Val Leu Val
 130 135 140
 5Trp Thr Asp Phe Thr Ala Cys Val Lys Ile Trp Cys Leu Lys Thr Gly
 145 150 155 160
 Arg Val Val Glu Ile Arg Asp Pro Lys Phe Pro Gly Lys Asp Gly Lys
 165 170 175
 Gly Trp Gly Tyr Arg Pro Ala Asp Asp Thr Gly Leu Arg Asn Gly Arg
 10 180 185 190
 Gly Gln Gly Arg Val Leu Ala Leu Leu Cys Arg Ala Ser Gly Thr Asp
 195 200 205
 Ile Leu Leu Leu Leu Ala Pro Gln Thr Tyr Lys Val Leu Asn Arg Val
 210 215 220
 15Glu Leu Pro Thr Thr Asp Ala Ala Gly Leu Arg Trp Ser Arg Asp Gly
 225 230 235 240
 Arg Trp Leu Ala Ile Trp Asp Ala Ala Ser Ala Gly Tyr Lys Leu
 245 250 255

20<210> 197

<211> 723

<212> DNA

<213> *Cochliobolus heterostrophus*

25<400> 197

atggacggtg gatgttgctg agtggccgat gaattcgccg atgaagatga cgtcgagtgg 60
 gagcgctgtg aagccgtgta caagtacacg actggttcgt catgcgcagt ttggataacg 120
 agacgattgg ggtcgcaggg gtgccagagg agtgctttca caggagcgta cattatgagg 180
 atcgagcggg gacggagact tcgcagggtcc caaatccaga ctgtgcaggg tgtgctgtcg 240
 30tctctgcttg cgcacattgt gccttcggag ttgaagctga gcatgccgat gccttgtttt 300
 aggagcgcat ttctgttctt ttctagggcg gctttgggag gtgtggcggg ttgtggtgtg 360
 agtgtgaaac tacgggcgcc caggttgtcg acttgctctg tgtacactgg tgcgctgggt 420
 acgtcgatga cgggtgtgtg gtcgaggaac aggatgggtg cgaatgtgctg ttagaaaagg 480
 atgcgaacac gacgggtcca gccgccaact gcgagacgtt catgtccagg gaccattct 540
 35agactcttga tgcctaggcc ttctacatcc cattcgctga cgtcctcgga tgcttcgcgg 600
 gttatggtgc ggtacaaatg cccatccgcc gtatatatca aagcttgtaa cccgcagacg 660
 cagcgtccca gatggccagc cagcgcccgt cagactcca tctcagacca gcggcgtctg 720
 tag 723

40<210> 198

130

<211> 240

<212> PRT

<213> Cochliobolus heterostrophus

5<400> 198

```

Met Asp Gly Gly Cys Cys Val Val Ala Asp Glu Phe Ala Asp Glu Asp
  1              5              10              15
Asp Val Glu Trp Glu Arg Cys Glu Ala Val Tyr Lys Tyr Thr Thr Gly
      20              25              30
10Ser Ser Cys Ala Val Trp Ile Thr Arg Arg Leu Gly Ser Gln Gly Cys
      35              40              45
Gln Arg Ser Ala Phe Thr Gly Ala Tyr Ile Met Arg Ile Glu Arg Gly
      50              55              60
Arg Arg Leu Arg Arg Ser Gln Ile Gln Thr Val Gln Gly Val Leu Ser
1565              70              75              80
Ser Leu Leu Ala His Ile Val Pro Ser Glu Leu Lys Leu Ser Met Pro
      85              90              95
Met Pro Cys Phe Arg Ser Ala Phe Ser Phe Phe Ser Arg Ala Ala Leu
      100              105              110
20Gly Gly Val Ala Gly Cys Gly Val Ser Val Lys Leu Arg Ala Pro Arg
      115              120              125
Leu Ser Thr Cys Ser Val Tyr Thr Gly Ala Leu Gly Thr Ser Met Thr
      130              135              140
Gly Val Trp Ser Arg Asn Arg Met Gly Ala Asn Val Arg Val Glu Arg
25145              150              155              160
Met Arg Thr Arg Arg Ser Gln Pro Pro Thr Ala Arg Arg Ser Cys Pro
      165              170              175
Gly Thr His Ser Arg Leu Leu Met Pro Arg Pro Ser Thr Ser His Ser
      180              185              190
30Leu Thr Ser Ser Asp Ala Ser Arg Val Met Val Arg Tyr Lys Cys Pro
      195              200              205
Ser Ala Val Tyr Ile Lys Ala Cys Asn Pro Gln Thr Gln Arg Pro Arg
      210              215              220
Trp Pro Ala Ser Ala Arg His Asp Ser Ile Ser Asp Gln Arg Arg Leu
35225              230              235              240

```

<210> 199

<211> 1647

<212> DNA

40<213> Cochliobolus heterostrophus

131

<400> 199

atgaacgtca	agcaagcggc	atgtctgaat	tgccgcacaaa	gcaagataaa	atgccggcgc	60
gaagaaggcg	cttctgtgtg	tgaagatgc	tctagcgtag	gcgtcgaatg	cattataccc	120
gagttccata	ttggtaggca	aaagggcgtg	aaaaacaaac	gatcaggggt	ggagaaagca	180
5atctaccaag	tagaagaagc	aatcaagaag	agaaaatcag	acgtagctgt	caaccagagc	240
acgttacagc	atgtgcaaca	gcttttgaac	gaagcacaa	gagacgttgg	ccctagtcaa	300
gatgcaaaat	caccgccagt	actagcagaa	ctatcttatg	tgccagcaaa	agaagttgcc	360
agcacttcaa	gcgatgatca	gcttgccgtt	gaagatgtcg	agaatccgct	tcagctttta	420
gcccgcgcat	cagacttgag	gattgccacc	acccacagct	cgtacaatac	aagtgtcgcc	480
10agcccagaag	gcaggtttac	tggtagcgag	caaagcgcat	tcctcgatgt	tcactacttc	540
ttcttaccac	tgaaggcgca	tttggaacaa	ggatctgggt	tagatccaat	tgatgtagga	600
ttggttacca	aagatgaagc	ggagatgctc	ctccaatatt	tccacaaaag	actagctcac	660
acgcgctggg	gtctagaccc	agtgggtgat	actctacctt	ttgtccgaaa	ccgctcagcc	720
tttctgttta	cgacattgct	ggctgtgacg	gccgtcttcc	taccagaaac	gtctgctttg	780
15gccaaaagac	tacttcttca	ccgcagggtt	ctagctgaac	aggtcattgt	tcgaaagtac	840
agatccggtg	aaatcgctct	ggcattcatg	gtgagcatac	catggatgcc	cccagggtcg	900
catgcaagcg	acgacgacac	aagtctctat	ctagctacgg	cattgtctat	ttctttggat	960
cttatgctag	acaaagtcac	cactccatct	acgtcctttg	gtccggagct	cacgaggcag	1020
atgcccaaag	cagagtgtct	tgacgcaaga	aaagcactag	ctatggatgg	tttcgaggac	1080
20attgacccga	cttctgaatg	gggccagcga	ctgcttcgtc	ggagagaaaag	ggtctggatt	1140
gcgctgtttg	tgctagagcg	tgccgtgtgc	ctcgctcggt	gccgcagcta	ctgtgtacca	1200
aagacgtgct	tgattcaata	cagcgataaa	tggcatgacc	accagcactc	ggatgcccag	1260
gacgggtccg	tagtatccat	ggcagtatta	cgtcgcgatc	tcgacaaact	ttttgccgaa	1320
gtacgcacgc	gatgcgacaa	ctatggctcg	gccgaagtag	gttcccaggt	tgccgcaggaa	1380
25atcgacaagt	caattgaggg	cttcttcgac	aattggtctc	gggcatggcc	ttcagttata	1440
agtgacccag	agagcaagag	cctacccctt	tatgtcgaga	tactcggttac	acacacacga	1500
ctctcgacct	actcaatgct	tctgaacctt	ccgagcgctc	caccagaagt	caagcgctcg	1560
ttccgcaagt	ctgcgttatc	ctcggcgctc	aatgttatgc	gccgcagcaa	tccaaggcga	1620
gggacctctc	aagtcaatgc	ccaataa				1647

30

<210> 200

<211> 548

<212> PRT

<213> Cochliobolus heterostrophus

35

<400> 200

Met Asn Val Lys Gln Ala Ala Cys Leu Asn Cys Arg Lys Ser Lys Ile

1

5

10

15

Lys Cys Arg Arg Glu Glu Gly Ala Ser Val Cys Glu Arg Cys Ser Ser

40

20

25

30

132

Val Gly Val Glu Cys Ile Ile Pro Glu Phe His Ile Gly Arg Gln Lys
 35 40 45
 Gly Val Lys Asn Lys Arg Ser Gly Leu Glu Lys Ala Ile Tyr Gln Val
 50 55 60
 5Glu Glu Ala Ile Lys Lys Arg Lys Ser Asp Val Ala Val Asn Gln Ser
 65 70 75 80
 Thr Leu Gln His Leu Gln Gln Leu Leu Asn Glu Ala Gln Gly Asp Val
 85 90 95
 Gly Pro Ser Gln Asp Ala Lys Ser Pro Pro Val Leu Ala Glu Leu Ser
 10 100 105 110
 Tyr Val Pro Ala Lys Glu Val Ala Ser Thr Ser Ser Asp Asp Gln Leu
 115 120 125
 Ala Val Glu Asp Val Glu Asn Pro Leu Gln Leu Leu Ala Arg Ala Ser
 130 135 140
 15Asp Leu Arg Ile Ala Thr Thr Pro Gln Ser Tyr Asn Thr Ser Val Ala
 145 150 155 160
 Ser Pro Glu Gly Arg Phe Thr Gly Ser Glu Gln Ser Ala Phe Leu Asp
 165 170 175
 Val His His Phe Phe Leu Pro Met Lys Ala His Leu Asp Gln Gly Ser
 20 180 185 190
 Gly Leu Asp Pro Ile Asp Val Gly Leu Val Thr Lys Asp Glu Ala Glu
 195 200 205
 Met Leu Leu Gln Tyr Phe His Lys Arg Leu Ala His Thr Arg Trp Gly
 210 215 220
 25Leu Asp Pro Val Val His Thr Leu Pro Phe Val Arg Asn Arg Ser Ala
 225 230 235 240
 Phe Leu Phe Thr Thr Leu Leu Ala Val Thr Ala Val Phe Leu Pro Glu
 245 250 255
 Thr Ser Ala Leu Ala Lys Arg Leu Leu Leu His Arg Arg Phe Leu Ala
 30 260 265 270
 Glu Gln Val Ile Val Arg Lys Tyr Arg Ser Val Glu Ile Val Leu Ala
 275 280 285
 Phe Met Val Ser Ile Pro Trp Met Pro Pro Gly Ser His Ala Ser Asp
 290 295 300
 35Asp Asp Thr Ser Leu Tyr Leu Ala Thr Ala Leu Ser Ile Ser Leu Asp
 305 310 315 320
 Leu Met Leu Asp Lys Val Ile Thr Pro Ser Thr Ser Phe Gly Pro Glu
 325 330 335
 Leu Thr Arg Gln Met Pro Lys Ala Glu Cys Leu Asp Ala Arg Lys Ala
 40 340 345 350

133

Leu Ala Met Asp Gly Phe Glu Asp Ile Asp Pro Thr Ser Glu Trp Gly
 355 360 365
 Gln Arg Leu Leu Arg Arg Arg Glu Arg Val Trp Ile Ala Leu Phe Val
 370 375 380
 5Leu Glu Arg Gly Val Cys Leu Ala Arg Gly Arg Ser Tyr Cys Val Pro
 385 390 395 400
 Lys Thr Cys Leu Ile Gln Tyr Ser Asp Lys Trp His Asp His Gln His
 405 410 415
 Ser Asp Ala Gln Asp Gly Pro Leu Val Ser Met Ala Val Leu Arg Arg
 10 420 425 430
 Asp Leu Asp Asn Leu Phe Ala Glu Val Arg Thr Arg Cys Asp Asn Tyr
 435 440 445
 Gly Ser Ala Glu Val Gly Ser Gln Val Ala Gln Glu Ile Asp Lys Ser
 450 455 460
 15Ile Glu Gly Phe Phe Asp Asn Trp Ser Arg Ala Trp Pro Ser Val Ile
 465 470 475 480
 Ser Asp Pro Glu Ser Lys Ser Leu Pro Pro Tyr Val Glu Ile Leu Val
 485 490 495
 Thr His Thr Arg Leu Ser Thr Tyr Ser Met Leu Leu Asn His Pro Ser
 20 500 505 510
 Ala Pro Pro Glu Val Lys Arg Ser Phe Arg Lys Ser Ala Leu Ser Ser
 515 520 525
 Ala Leu Asn Val Met Arg Arg Ser Asn Pro Arg Arg Gly Thr Ser Gln
 530 535 540
 25Val Asn Ala Gln
 545

<210> 201

<211> 2271

30<212> DNA

<213> Cochliobolus heterostrophus

<400> 201

atggcgagcg	cagagcagac	aatcaacctc	aaggctcttt	cgccttcagc	ggaactagag	60
35ggcgcatca	ccctcgcg	ggtacctcgt	tctatcacgg	tcaaagagct	ccgcacccgc	120
atacacgatg	ctgtgcctc	caagcctgcc	cccgagcgca	tgcgctcat	atacagaggc	180
cgagtggtag	cgaatgatgc	agacactctg	actaccgtgt	ttggcgctga	caatatacgt	240
gagaacaaga	accaaagcct	tcacctcgtc	atacgagagc	tgctccaac	tgcatcttcg	300
cctgtcccgc	aatcgtcttc	tgtccacca	aacctcttcc	gctctgctgg	tccagatggc	360
40ccagccgcga	gccctctgca	gacgaatcca	tttcgggcta	taccacagac	acgaccggct	420

134

```

tcacaacctc aaatacccca gtcgcacctt ccgcctcatc gccttccggg acaagtgaac 480
cccatgccca taccattacc cgcacaactc catcaaacgt ttgctcaagc aatggcacac 540
caaggacaac aggggtgatga acagccctca gatcgaacta gcgagcagcc agatcaaggt 600
acaccggcag cgggggatag gacgcataca ccaatccctt caggaccgtc gaaccctcct 660
5ggaaatggcg accaggcgat caggcgagaa ggtggtgctc ctaatggagc acgatggaca 720
gttacggcct tcaatccact taacatagct gcgcgactcc cgccgctgt cgtcacattc 780
cctgtcccg c atgcactaac tttcggctgt ccgcgccttt ctagcgacaa ccagcggtta 840
ttgcctcgtg tgcacaggat cttcttggag acaaaacggg agattgataa cattcgagca 900
ttggtgcaac tgcctgggtg atctgatgca cagagtggag ggctcctcac ctgagatata 960
10cctgcctcgt tgaatatccc tgtatggcga atcgagcgac tacgtcagca cctgaacaca 1020
gtcaatcaaa atctggatgt cgttgaccgg gctctggcgt tgcttcctac agagcctgaa 1080
gtgacggcgc tcaggcgctc agctaccgag ttgaggggtg atgctgcgga attgagtatt 1140
gtgctcgatc gtcaacaggg cgaaacggcc agggctactt cggatacagc accaggggtg 1200
cccaccatag ctgcggcatc atcaactaca tcccagaccc gaccaggaga tgtgacacag 1260
15actgtaccga cagatgcacc tgcagagctg ttccttttgt caagtcccca gggctccggt 1320
ggagtctctc tcgatcagcg aggcacatac accacagccc caatggtgcc cactctacca 1380
ttccagagct tctcgagtca atttgcacag aacagacagc tcattgctgg tcttgggcag 1440
caaatggcac agggggacaaa ccacctgcat aatcaagtat ctaacatgca gccaacacca 1500
atagggcagc cagtagctgt tggacaggct caagatcata accgaggata tgatcagaat 1560
20cagaatcaga atcagaatca aaaccagaac cagaatgata atcagaatgg agtgcagcca 1620
gaagaaaatg atcggatggc caatatcgcc ggacatttgt ggctgatctt caagctcgct 1680
gtcttcgtct acgtcttcgc tggaggtggg ggtatttaca ggctgtaat gctaggtgct 1740
attgctggga ttgtctatct ggcacagatc ggcattgttg aggatcagat caactacgtg 1800
cgtcgccatt ttgaggctct tcttcctggt ggcgctatgg ccgaacgcgc tgcacaacct 1860
25atcaaccagc gccacagagg taacatatcg cccgaggaag cagcaaggcg aataactaaa 1920
caaagacaag aacaaagggt cgcctgggta cgcgagagct tgcgtggagt cgagcgcgct 1980
ttcactctct tcattgccag tctattccct ggtgtaggcg agagaatggt tcacgcacag 2040
gaagagagag agagactgga gagggtagca gcacgggaag agagagagag acaggaggag 2100
gaagcgagga agcgagaaga agacgccagg gcacagcagc aacagcagac cgatgagaaa 2160
30gctagtgaag ccaggggtga gatggacagt gaggttactc caagcagcag ttcaaagggc 2220
aaggagaggg ctgaggagca acacgttgat gggtcagcct catcttcatg a 2271

```

<210> 202

<211> 756

35<212> PRT

<213> Cochliobolus heterostrophus

<400> 202

Met Ala Asp Ala Glu Gln Thr Ile Asn Leu Lys Val Leu Ser Pro Ser

40 1

5

10

15

135

Ala	Glu	Leu	Glu	Gly	Gly	Ile	Thr	Leu	Ala	Gly	Leu	Pro	Ala	Ser	Ile
		20						25					30		
Thr	Val	Lys	Glu	Leu	Arg	Thr	Arg	Ile	His	Asp	Ala	Val	Pro	Ser	Lys
		35					40					45			
5Pro	Ala	Pro	Glu	Arg	Met	Arg	Leu	Ile	Tyr	Arg	Gly	Arg	Val	Val	Ala
		50				55					60				
Asn	Asp	Ala	Asp	Thr	Leu	Thr	Thr	Val	Phe	Gly	Ala	Asp	Asn	Ile	Arg
65					70					75				80	
Glu	Asn	Lys	Asn	Gln	Ser	Leu	His	Leu	Val	Ile	Arg	Glu	Leu	Pro	Pro
10				85					90					95	
Thr	Ala	Ser	Ser	Pro	Val	Pro	Gln	Ser	Ser	Ser	Val	Pro	Pro	Asn	Leu
			100				105					110			
Phe	Arg	Ser	Ala	Gly	Pro	Asp	Gly	Pro	Ala	Ala	Ser	Pro	Leu	Gln	Thr
		115					120					125			
15Asn	Pro	Phe	Arg	Ala	Ile	Pro	Gln	Thr	Arg	Pro	Ala	Ser	Gln	Pro	Gln
		130					135				140				
Ile	Pro	Gln	Ser	His	Leu	Pro	Pro	His	Arg	Leu	Pro	Gly	Gln	Val	Asn
145					150					155				160	
Pro	Ile	Pro	Ile	Pro	Leu	Pro	Ala	Gln	Leu	His	Gln	Thr	Phe	Ala	Gln
20				165					170					175	
Ala	Met	Ala	His	Gln	Gly	Gln	Gln	Gly	Asp	Glu	Gln	Pro	Ser	Asp	Arg
			180					185				190			
Thr	Ser	Glu	Gln	Pro	Asp	Gln	Gly	Thr	Pro	Ala	Ala	Gly	Asp	Arg	Thr
		195					200					205			
25His	Thr	Pro	Ile	Pro	Ser	Gly	Pro	Ser	Asn	Pro	Pro	Gly	Asn	Gly	Asp
		210				215						220			
Gln	Ala	Ile	Arg	Arg	Glu	Gly	Val	Ala	Pro	Asn	Gly	Ala	Arg	Trp	Thr
225					230					235				240	
Val	Thr	Ala	Phe	Asn	Pro	Leu	Asn	Ile	Ala	Ala	Arg	Leu	Pro	Pro	Pro
30				245					250					255	
Val	Val	Thr	Phe	Pro	Val	Pro	His	Ala	Leu	Thr	Phe	Gly	Arg	Pro	Pro
			260					265				270			
Leu	Ser	Ser	Asp	Asn	Gln	Arg	Leu	Leu	Pro	Arg	Val	His	Arg	Ile	Phe
		275					280					285			
35Leu	Glu	Thr	Lys	Arg	Glu	Ile	Asp	Asn	Ile	Arg	Ala	Leu	Leu	Gln	Leu
		290					295				300				
Pro	Gly	Ala	Ser	Asp	Ala	Gln	Ser	Gly	Gly	Leu	Leu	Thr	Ser	Asp	Ile
305					310					315				320	
Pro	Ala	Ser	Leu	Asn	Ile	Pro	Val	Trp	Arg	Ile	Glu	Arg	Leu	Arg	Gln
40				325					330					335	

136

His Leu Asn Thr Val Asn Gln Asn Leu Asp Val Val Asp Arg Ala Leu
 340 345 350
 Ala Leu Leu Pro Thr Glu Pro Glu Val Thr Ala Leu Arg Arg Ser Ala
 355 360 365
 5Thr Glu Leu Arg Val Asp Ala Ala Glu Leu Ser Ile Val Leu Asp Arg
 370 375 380
 Gln Gln Gly Glu Thr Ala Arg Ala Thr Ser Asp Thr Ala Pro Gly Val
 385 390 395 400
 Pro Thr Ile Ala Ala Ala Ser Ser Thr Thr Ser Gln Thr Arg Pro Gly
 10 405 410 415
 Asp Val Thr Gln Thr Val Pro Thr Asp Ala Pro Ala Glu Leu Phe Leu
 420 425 430
 Leu Ser Ser Pro Gln Gly Pro Val Gly Val Leu Phe Asp Gln Arg Gly
 435 440 445
 15Thr Tyr Thr Thr Ala Pro Met Val Pro Thr Leu Pro Phe Gln Ser Phe
 450 455 460
 Ser Ser Gln Phe Ala Gln Asn Arg Gln Leu Ile Ala Gly Leu Gly Gln
 465 470 475 480
 Gln Met Ala Gln Gly Thr Asn His Leu His Asn Gln Val Ser Asn Met
 20 485 490 495
 Gln Pro Thr Pro Ile Gly Gln Pro Val Ala Val Gly Gln Ala Gln Asp
 500 505 510
 His Asn Arg Gly Tyr Asp Gln Asn Gln Asn Gln Asn Gln Asn Gln Asn
 515 520 525
 25Gln Asn Gln Asn Asp Asn Gln Asn Gly Val Gln Pro Glu Glu Asn Asp
 530 535 540
 Arg Met Ala Asn Ile Ala Gly His Leu Trp Leu Ile Phe Lys Leu Ala
 545 550 555 560
 Val Phe Val Tyr Val Phe Ala Gly Gly Gly Gly Ile Tyr Arg Pro Val
 30 565 570 575
 Met Leu Gly Ala Ile Ala Gly Ile Val Tyr Leu Ala Gln Ile Gly Met
 580 585 590
 Phe Glu Asp Gln Ile Asn Tyr Val Arg Arg His Phe Glu Ala Leu Leu
 595 600 605
 35Pro Val Gly Ala Met Ala Glu Arg Ala Ala Gln Pro Ile Asn Gln Arg
 610 615 620
 Pro Arg Gly Asn Ile Ser Pro Glu Glu Ala Ala Arg Arg Ile Leu Gln
 625 630 635 640
 Gln Arg Gln Glu Gln Arg Phe Ala Trp Leu Arg Glu Ser Leu Arg Gly
 40 645 650 655

137

Val Glu Arg Ala Phe Thr Leu Phe Ile Ala Ser Leu Phe Pro Gly Val
660 665 670
Gly Glu Arg Met Val His Ala Gln Glu Glu Arg Glu Arg Leu Glu Arg
675 680 685
5Val Ala Ala Arg Glu Glu Arg Glu Arg Gln Glu Glu Glu Ala Arg Lys
690 695 700
Arg Glu Glu Asp Ala Arg Ala Gln Gln Gln Gln Gln Thr Asp Glu Lys
705 710 715 720
Ala Ser Glu Ala Arg Val Glu Met Asp Ser Glu Val Thr Pro Ser Ser
10 725 730 735
Ser Ser Lys Gly Lys Glu Arg Ala Glu Glu Gln His Val Asp Gly Ser
740 745 750
Ala Ser Ser Ser
755

15

<210> 203
<211> 489
<212> DNA
<213> Cochliobolus heterostrophus

20

<400> 203
atggcgctaa tccttcccca gtgggtacgt catggggttag ccgctcatgt ggattttccc 60
acaccaccca acgccttcgc cgtcatcttc ccgctttgcc attgcgaatc tcctcatctt 120
gatgtggcgc agaaaacaac ggtagaaatc gcaaatgcat gcatcataac atggcgactc 180
25cctgttcgcg ccagctcatt cgagcttcta tccgaccgcg atgcaatata cccacacacc 240
cacatgcgcg cctctctgcc ccatctcgca ccttcacgtc gactcgagcg tcatatagcg 300
aacaaaattt tcacaggaag gagtctttta ggtcgcggct caattccgca ctcaagaata 360
ccaaagtcaa atgggagcca ataccaattg cactcggtat tggcttcctg ggtgcatttc 420
agctatatcg catacaacgc agagaaaagc atacagaagc cgagagaagg gatgcggatg 480
30gcaatgtag 489

<210> 204
<211> 162
<212> PRT

35<213> Cochliobolus heterostrophus

<400> 204

Met Ala Leu Ile Pro Pro Gln Trp Val Arg His Gly Leu Ala Ala His
1 5 10 15
40Val Asp Phe Pro Thr Pro Pro Asn Ala Phe Ala Val Ile Phe Pro Leu

138

	20		25		30	
	Cys His Cys Glu Ser Pro His Leu Asp Val Ala Glu Lys Thr Thr Val					
	35		40		45	
	Glu Ile Ala Asn Ala Cys Ile Ile Thr Trp Arg Leu Pro Val Arg Ala					
5	50		55		60	
	Ser Ser Phe Glu Leu Leu Ser Asp Arg Asp Ala Ile Tyr Pro His Thr					
	65		70		75	
	His Met Arg Ala Ser Leu Pro His Leu Ala Pro Ser Arg Arg Leu Glu					
	85		90		95	
10	Arg His Ile Ala Asn Lys Ile Phe Thr Gly Arg Ser Leu Leu Gly Arg					
	100		105		110	
	Gly Ser Ile Pro His Ser Arg Ile Pro Lys Ser Asn Gly Ser Gln Tyr					
	115		120		125	
	Gln Leu His Ser Val Leu Ala Ser Trp Val His Phe Ser Tyr Ile Ala					
15	130		135		140	
	Tyr Asn Ala Glu Lys Ser Ile Gln Lys Pro Arg Glu Gly Met Arg Met					
	145		150		155	
	Ala Met					

20

<210> 205

<211> 1581

<212> DNA

<213> Cochliobolus heterostrophus

25

<220>

<221> misc_feature

<222> (1)...(1581)

<223> n = any nucleotide

30

<400> 205

atgcatcata	acatggcgac	tcctgttcg	cgccagctca	ttcgagcttc	tatccgaccg	60
cgatgcaata	taccacacaca	cccacatgog	cgctctctcg	ccccatctcg	caccttcacg	120
tcgactcgag	cgtcatatag	cgaacaaaat	tttcacagga	aggagtcttt	taggtcgcg	180
35ctcaattccg	cactcaagaa	taccaaagtc	aaatgggagc	caataccaat	tgactcggt	240
attggcttcc	tgggtgcatt	tcagctatat	cgcatacaac	gcagagaaaa	gcatacagaa	300
gccgagagaa	gggatgcgga	tggcaatgta	gtggatcagc	aaggctgtcc	gaagaagcgc	360
gaaagaataa	gaccgagcgg	accatggacc	gttcagggtca	tgtctaccct	tcctctcaag	420
gcgttgctcg	gactgtgggg	tcgcttcaat	gagatcgaca	taccctacta	ccttcacta	480
40catgtatacc	ccaacctcgc	cgcctttttc	taccgcaccc	tcaaaccg	tgtacgtcct	540

139

```

ctagatccca accccaacgc agtactctct cccgcagacg gcaagatcat tcaatttggc      600
accatcgagc acggcgaagt tgagcaagtc aaaggtgtaa catatagttt ggacgctctg      660
ctaggatcta caaggncag tacaccagag caaaatgtag caaattccca aattcgcgct      720
agtgagcacg agaagacacc acaagacgaa gaggacactg tgcgcgcgga tgaggaattt      780
5gcaaacgtga acggtatctc atatactcta ccaaactct tctccggacc atggccaaaa      840
gacgggaagc ctgctgaaat gccgacggat caatcagttc cgtcaaagcc atcgtcagaa      900
gccgaagtac gtgccgacct tgccttgagt gaatcacagc gcccatggtg ggcacccgcc      960
tcattaaaga cacctacggg tctctactac tgcgttgtag atcttgcgcc aggcgactac     1020
cacagggttcc actcacctgt atcatgggtt gttgagtcgc gtcgtaactt tgctggcgag     1080
10ctttatagtg tatcgcccta cctacaacgc actatgcctg gtctctttac cctgaacgag     1140
cgtgtggttc tcctaggaag atggcgctgg ggtttctttt cctacactcc ggtcggcgca     1200
accaacgttg gttccattaa gatcaacttt gatcgcgaaac ttgcacaaa cagcttaaca     1260
accgacactg cggcggaccg tgctgcgga gaagccgctg cccgtggtga gccgtattct     1320
ggattcgctg aggcctccta cagcagcgca agccgtgtct tgggagggta cgcactcaag     1380
15cgcggcgagg aaatgggtgg ttttcagttg ggcagtacaa ttgtcttagt ctttgaagcg     1440
ccgaagggca ttcgacctag tttggacgag ggcttttagtg gtacacgtgg cgagagaaaa     1500
gggtgggtttc actggaatat cgaacaaggg caaaaagtca aggttggcga ggcgttgggt     1560
tatgttgaag aagttcagta a                                     1581

```

20<210> 206

<211> 526

<212> PRT

<213> *Cochliobolus heterostrophus*

25<220>

<221> SITE

<222> (1)...(526)

<223> Xaa = any amino acid

30<400> 206

```

Met His His Asn Met Ala Thr Pro Cys Ser Arg Gln Leu Ile Arg Ala
 1              5              10              15
Ser Ile Arg Pro Arg Cys Asn Ile Pro Thr His Pro His Ala Arg Leu
          20              25              30
35Ser Ala Pro Ser Arg Thr Phe Thr Ser Thr Arg Ala Ser Tyr Ser Glu
          35              40              45
Gln Asn Phe His Arg Lys Glu Ser Phe Arg Ser Arg Leu Asn Ser Ala
          50              55              60
Leu Lys Asn Thr Lys Val Lys Trp Glu Pro Ile Pro Ile Ala Leu Gly
4065              70              75              80

```

Ile	Gly	Phe	Leu	Gly	Ala	Phe	Gln	Leu	Tyr	Arg	Ile	Gln	Arg	Arg	Glu		
				85							90			95			
Lys	His	Thr	Glu	Ala	Glu	Arg	Arg	Asp	Ala	Asp	Gly	Asn	Val	Val	Asp		
				100							105			110			
5Gln	Gln	Gly	Arg	Pro	Lys	Lys	Arg	Glu	Arg	Ile	Arg	Pro	Ser	Gly	Pro		
				115							120			125			
Trp	Thr	Val	Gln	Val	Met	Ser	Thr	Leu	Pro	Leu	Lys	Ala	Leu	Ser	Arg		
				130							135			140			
Leu	Trp	Gly	Arg	Phe	Asn	Glu	Ile	Asp	Ile	Pro	Tyr	Tyr	Leu	His	Leu		
10145					150							155			160		
His	Val	Tyr	Pro	Asn	Leu	Ala	Ala	Phe	Phe	Tyr	Arg	Thr	Leu	Lys	Pro		
				165							170			175			
Gly	Val	Arg	Pro	Leu	Asp	Pro	Asn	Pro	Asn	Ala	Val	Leu	Ser	Pro	Ala		
				180							185			190			
15Asp	Gly	Lys	Ile	Ile	Gln	Phe	Gly	Thr	Ile	Glu	His	Gly	Glu	Val	Glu		
				195							200			205			
Gln	Val	Lys	Gly	Val	Thr	Tyr	Ser	Leu	Asp	Ala	Leu	Leu	Gly	Ser	Thr		
				210							215			220			
Arg	Xaa	Ser	Thr	Pro	Glu	Gln	Asn	Val	Ala	Asn	Ser	Gln	Ile	Arg	Ala		
20225					230							235			240		
Ser	Glu	His	Glu	Lys	Thr	Pro	Gln	Asp	Glu	Glu	Asp	Thr	Val	Arg	Ala		
				245							250			255			
Asp	Glu	Glu	Phe	Ala	Asn	Val	Asn	Gly	Ile	Ser	Tyr	Thr	Leu	Pro	Asn		
				260							265			270			
25Leu	Phe	Ser	Gly	Pro	Trp	Pro	Lys	Asp	Gly	Lys	Pro	Ala	Glu	Met	Pro		
				275							280			285			
Thr	Asp	Gln	Ser	Val	Pro	Ser	Lys	Pro	Ser	Ser	Glu	Ala	Glu	Val	Arg		
				290							295			300			
Ala	Asp	Leu	Ala	Leu	Ser	Glu	Ser	Gln	Arg	Pro	Trp	Trp	Ala	Pro	Ala		
30305					310							315			320		
Ser	Leu	Lys	Thr	Pro	Thr	Val	Leu	Tyr	Tyr	Cys	Val	Val	Tyr	Leu	Ala		
				325							330			335			
Pro	Gly	Asp	Tyr	His	Arg	Phe	His	Ser	Pro	Val	Ser	Trp	Val	Val	Glu		
				340							345			350			
35Ser	Arg	Arg	His	Phe	Ala	Gly	Glu	Leu	Tyr	Ser	Val	Ser	Pro	Tyr	Leu		
				355							360			365			
Gln	Arg	Thr	Met	Pro	Gly	Leu	Phe	Thr	Leu	Asn	Glu	Arg	Val	Val	Leu		
				370							375			380			
Leu	Gly	Arg	Trp	Arg	Trp	Gly	Phe	Phe	Ser	Tyr	Thr	Pro	Val	Gly	Ala		
40385					390							395			400		

141

Thr Asn Val Gly Ser Ile Lys Ile Asn Phe Asp Arg Glu Leu Arg Thr
 405 410 415
 Asn Ser Leu Thr Thr Asp Thr Ala Ala Asp Arg Ala Ala Glu Glu Ala
 420 425 430
 5Ala Ala Arg Gly Glu Pro Tyr Ser Gly Phe Ala Glu Ala Ser Tyr Thr
 435 440 445
 Ser Ala Ser Arg Val Leu Gly Gly Tyr Ala Leu Lys Arg Gly Glu Glu
 450 455 460
 Met Gly Gly Phe Gln Leu Gly Ser Thr Ile Val Leu Val Phe Glu Ala
 10465 470 475 480
 Pro Lys Gly Ile Arg Pro Ser Leu Asp Glu Gly Phe Ser Gly Thr Arg
 485 490 495
 Gly Glu Arg Lys Gly Gly Phe His Trp Asn Ile Glu Gln Gly Gln Lys
 500 505 510
 15Val Lys Val Gly Glu Ala Leu Gly Tyr Val Glu Glu Val Gln
 515 520 525

<210> 207

<211> 366

20<212> DNA

<213> Cochliobolus heterostrophus

<400> 207

atgcccgaga ccaattgtcg aagaagccct caattgactt gtcgatttcc tgcgcaacct 60
 25gggaacctac ttcggccgag ccatagttgt cgcacgcgt gcgtacttcg gcaaaaaggt 120
 tgtcgagatc gcgacgtaat actgccatgg atactagcgg accgtcctgg gcatccgagt 180
 gctggtggtc atgccattta tcgctgtatt gaatcaagca cgtctttggt acacagtagc 240
 tgcggccacg agcgaggcac acgccacgct ctagcacaaa cagcgcaatc cagacccttt 300
 ctctccgacg aagcagtcgc tggccccatt cagaagtcgg gtcaatgtcc tcgaaacct 360
 30ccatag 366

<210> 208

<211> 121

<212> PRT

35<213> Cochliobolus heterostrophus

<400> 208

Met Pro Glu Thr Asn Cys Arg Arg Ser Pro Gln Leu Thr Cys Arg Phe
 1 5 10 15
 40Pro Ala Gln Pro Gly Asn Leu Leu Arg Pro Ser His Ser Cys Arg Ile

142

	20		25		30
	Ala Cys Val Leu Arg Gln Lys Gly Cys Arg Asp Arg Asp Val Ile Leu				
	35		40		45
	Pro Trp Ile Leu Ala Asp Arg Pro Gly His Pro Ser Ala Gly Gly His				
5	50		55		60
	Ala Ile Tyr Arg Cys Ile Glu Ser Ser Thr Ser Leu Val His Ser Ser				
	65		70		75
	Cys Gly His Glu Arg Gly Thr Arg His Ala Leu Ala Gln Thr Ala Gln				
	85		90		95
10	Ser Arg Pro Phe Leu Ser Asp Glu Ala Val Ala Gly Pro Ile Gln Lys				
	100		105		110
	Ser Gly Gln Cys Pro Arg Asn His Pro				
	115		120		

15<210> 209

<211> 714

<212> DNA

<213> Cochliobolus heterostrophus

20<400> 209

atgtgcgcaa	gcagagacga	cagcacaccc	tgcacagtct	ggattttggga	cctgcgaagt	60
ctccgtcccc	gctcgatcct	cataatgtac	gctcctgtga	aagcactcct	ctggcacccc	120
tgcgacccca	atcgtctcgt	tatccaaact	gcgcatgacg	aaccagtcgt	gtacttgtac	180
acggcttcac	agcgcctcca	ctcgacgtca	tcttcatcgg	cgaattcatc	ggccactacg	240
25caacatccac	cgtccatcct	ctcgttatcc	cctcacattg	ctaaaccgcg	agtcgcgact	300
cccgcacgct	ggaccgtatc	ctggctctcg	ggccctgcag	ctagtagtaa	aaaaccttgt	360
ttcgcgctcg	cgcataactca	agcttccggt	gtcgtttggc	cggagggcaa	agaccagatt	420
ctgcggtttg	atcatgaaga	cgaagaagag	ggatgaagag	agggcggaaga	ggagggcgag	480
gaagcgggat	cagatgatag	tttgtatgat	atactgactg	gccggacacc	ggtaccgggt	540
30acaagagata	gcatggagga	aggggggttt	ggggatagta	caggcacagt	gcaggagcta	600
gatgatacgt	ttcgatcgcg	gcgacatgca	catcaagaac	atggaggaca	tgaggaacac	660
gagtactttg	aagaaggtgt	gttgggagat	agtggcatga	gtgaaatgtt	ttga	714

<210> 210

35<211> 237

<212> PRT

<213> Cochliobolus heterostrophus

<400> 210

40Met Cys Ala Ser Arg Asp Asp Ser Thr Pro Cys Thr Val Trp Ile Trp

143

1	5	10	15												
Asp	Leu	Arg	Ser	Leu	Arg	Pro	Arg	Ser	Ile	Leu	Ile	Met	Tyr	Ala	Pro
	20		25		30										
Val	Lys	Ala	Leu	Leu	Trp	His	Pro	Cys	Asp	Pro	Asn	Arg	Leu	Val	Ile
5	35		40		45										
Gln	Thr	Ala	His	Asp	Glu	Pro	Val	Val	Tyr	Leu	Tyr	Thr	Ala	Ser	Gln
	50		55		60										
Arg	Ser	His	Ser	Thr	Ser	Ser	Ser	Ser	Ala	Asn	Ser	Ser	Ala	Thr	Thr
65			70		75										80
10Gln	His	Pro	Pro	Ser	Ile	Leu	Ser	Leu	Ser	Pro	His	Ile	Ala	Lys	Pro
			85		90										95
Ala	Val	Ala	Thr	Pro	Ala	Arg	Trp	Thr	Val	Ser	Trp	Leu	Ser	Gly	Pro
			100		105										110
Ala	Ala	Ser	Ser	Lys	Lys	Pro	Cys	Phe	Ala	Leu	Ala	His	Thr	Gln	Ala
15	115		120		125										
Ser	Val	Val	Val	Trp	Pro	Glu	Gly	Lys	Asp	Gln	Ile	Leu	Arg	Phe	Asp
	130		135		140										
His	Glu	Asp	Glu	Glu	Glu	Gly	Glu	Glu	Glu	Gly	Glu	Glu	Glu	Gly	Glu
145			150		155										160
20Glu	Ala	Gly	Ser	Asp	Asp	Ser	Leu	Tyr	Asp	Ile	Leu	Thr	Gly	Arg	Thr
			165		170										175
Pro	Val	Pro	Gly	Thr	Arg	Asp	Ser	Met	Glu	Glu	Gly	Gly	Phe	Gly	Asp
			180		185										190
Ser	Thr	Gly	Thr	Val	Gln	Glu	Leu	Asp	Asp	Thr	Phe	Arg	Ser	Arg	Arg
25	195		200		205										
His	Ala	His	Gln	Glu	His	Gly	Gly	His	Glu	Glu	His	Glu	Tyr	Phe	Glu
	210		215		220										
Glu	Gly	Val	Leu	Gly	Asp	Ser	Gly	Met	Ser	Glu	Met	Phe			
225			230		235										

30

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 2002/042444 A3

(51) International Patent Classification⁷: **C12N 15/52**,
15/53, 15/31, 1/21, 15/82, 9/02, 9/00, C07K 14/37, C12Q
1/18, G01N 33/573, 33/68, A61K 35/00, A01N 61/00,
A01H 5/00

(21) International Application Number:
PCT/US2001/043381

(22) International Filing Date:
21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,649 22 November 2000 (22.11.2000) US
60/252,732 22 November 2000 (22.11.2000) US

(71) Applicants (*for all designated States except US*): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldalee 215, CH-4058 (CH). **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; 20 Fomwood Drive, Suite 105, Ithaca, NY 14850 (US).

(71) Applicants and

(72) Inventors: **YODER, Olen** [US/US]; 4939 Concannon Court, San Diego, CA 91230 (US). **TURGEON, Barbara, G.** [CA/US]; 4939 Concannon Court, San Diego, CA 92130 (US). **LU, Shen-wen** [US/US]; 604 Winston Court, Apt. 4, Ithaca, NY 14850 (US).

(74) Agents: **KERNER, Ann-Louise** et al.; Hale and Dorr LLP, 60 State Street, Boston, MA 02109 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
31 December 2003

(15) Information about Correction:

Previous Correction:

see PCT Gazette No. 14/2003 of 3 April 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: FUNGAL GENE CLUSTER ASSOCIATED WITH PATHOGENESIS

(57) Abstract: Methods to identify orthologs ofungal CPS1 genes as well as fungal iron reductase and permease/and or MFS transporter genes, and uses thereof are provided.



WO 2002/042444 A3

INTERNATIONAL SEARCH REPORT

Inter I Application No

PCT/US 01/43381

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7	C12N15/52	C12N15/53	C12N15/31	C12N1/21	C12N15/82
	C12N9/02	C12N9/00	C07K14/37	C12Q1/18	G01N33/573
	G01N33/68	A61K35/00	A01N61/00	A01H5/00	

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>DATABASE EMBL 'Online! 25 January 2001 (2001-01-25) LU S.W.ET AL: "Cochliobolus heterostrophus peptide synthetase-like protein (CPS1) gene, complete cds" retrieved from EMBL Database accession no. AF332878 XP002230678 the whole document</p> <p style="text-align: center;">---</p>	1-18, 21, 27-52, 54, 56-62
A	<p>TURGAY K ET AL: "A GENERAL APPROACH FOR IDENTIFYING AND CLONING PEPTIDE SYNTHETASE GENES" PEPTIDE RESEARCH, NATICK, MA, US, vol. 7, no. 5, September 1994 (1994-09), pages 238-241, XP000978191 ISSN: 1040-5704</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- * & * document member of the same patent family

Date of the actual completion of the international search

14 February 2003

Date of mailing of the international search report

05.06.03

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Turri, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/43381

G.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>NIKOLSKAYA A N ET AL: "Identification of peptide synthetase-encoding genes from filamentous fungi producing host-selective phytotoxins or analogs" GENE, ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, NL, vol. 165, no. 2, 20 November 1995 (1995-11-20), pages 207-211, XP004043143 ISSN: 0378-1119</p> <p>---</p>	
A	<p>MAIER FRANK J ET AL: "Mutagenesis via insertional - or restriction enzyme-mediated - integration (REMI) as a tool to tag pathogenicity related genes in plant pathogenic fungi." BIOLOGICAL CHEMISTRY, vol. 380, no. 7-8, July 1999 (1999-07), pages 855-864, XP009005657 ISSN: 1431-6730</p> <p>---</p>	
A	<p>LU SHUNWEN ET AL: "Tagged mutations at the Tox1 locus of Cochliobolus heterostrophus by restriction enzyme-mediated integration." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 91, no. 26, 1994, pages 12649-12653, XP002230675 1994 ISSN: 0027-8424</p> <p>---</p>	
T	<p>LU S.W. ET AL: "A gene cluster from the corn Cochliobolus heterostrophus required for nonribosomal peptide biosynthesis and general virulence of fungi" SECONDARY METABOLISM AND PATHOGENICITY ABSTRACTS, ABSTRACTS NUMBERS 197-287, 'Online! 14 December 2000 (2000-12-14), XP002230676 Retrieved from the Internet: <URL:http://www.fgsc.net/asilo99/posterabs5.htm> 'retrieved on 2003-02-10! cited in the application Abstracts of the Twentieth Fungal Genetics Conference, held March 23-28, 1999, Pacific Grove, California. See Abstract no. 245</p> <p>---</p> <p style="text-align: center;">-/--</p>	

INTERNATIONAL SEARCH REPORT

Inter Application No

PCT/US 01/43381

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	<p>LU S -W ET AL: "Cochliobolus heterostrophus and Fusarium graminearum: Evidence for a common virulence factor." PHYTOPATHOLOGY, vol. 91, no. 6 Supplement, June 2001 (2001-06), page S56 XP001122359 Joint Meeting of the American Phytopathological Society, the Mycological Society of America, and the Society of Nematologists; Salt Lake City, Utah, USA; August 25-29, 2001 ISSN: 0031-949X</p>	
A	<p>STACHELHAUS ET AL: "Modular structure of peptide synthetases" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 270, no. 11, 17 March 1995 (1995-03-17), pages 6163-6169, XP002113293 ISSN: 0021-9258</p>	
A	<p>STACHELHAUS AND M A MARAHIEL T: "Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis" FEMS MICROBIOLOGY LETTERS, AMSTERDAM, NL, vol. 125, 1995, pages 3-14, XP002094834 ISSN: 0378-1097</p>	
A	<p>TURGAY K ET AL: "FOUR HOMOLOGOUS DOMAINS IN THE PRIMARY STRUCTURE OF GRSB ARE RELATED TO DOMAINS IN A SUPERFAMILY OF ADENYLATE-FORMING ENZYMES" MOLECULAR MICROBIOLOGY, BLACKWELL SCIENTIFIC, OXFORD, GB, vol. 6, no. 4, 1992, pages 529-546, XP001055896 ISSN: 0950-382X</p>	
A	<p>YODER O C ET AL: "Fungal genomics and pathogenicity." CURRENT OPINION IN PLANT BIOLOGY, vol. 4, no. 4, August 2001 (2001-08), pages 315-321, XP002230677 ISSN: 1369-5266</p>	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/43381

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 53, 55
because they relate to subject matter not required to be searched by this Authority, namely:
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☒ Claims Nos.: 19, 20, 22-26
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4 (all partially), 5, 6, 7 (partially), 8-18, 21, 27 (partially)-52, 54, 56-62

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-4 (all partially), 5, 6, 7 (partially), 8-18, 21, 27 (partially)-52, 54, 56-62

Isolated polynucleotides comprising the open reading frame of SEQ ID NO:46 and encoding a polypeptide having at least 80% identity to SEQ ID NO:47.

Isolated polypeptides; expression cassettes and vectors comprising said polynucleotides. Host cells.

Methods for identifying inhibitors of the polypeptide and for identifying agents that alter the phenotype of a fungal pathogen.

Isolated antibodies.

Therapeutic methods.

Transformed plants.

2. Claims: 1-4 (all partially), 5, 6, 7 (partially), 8-18, 21, 27 (partially)-52, 54, 56-62

As group 1, but with SEQ ID NOs:48 and 49

3. Claims: 1-4 (all partially), 5, 6, 7 (partially), 8-18, 21, 27 (partially)-52, 54, 56-62

As group 1, but with SEQ ID NOs:55 and 56

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 48-51, and 56-59 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.1

Claims Nos.: 53, 55

Rule 39.1(v) PCT - Presentation of information

Continuation of Box I.2

Claims Nos.: 19, 20, 22-26

Present claims 19 and 25 relate to an extremely large number of possible compounds binding to the CPS1 polypeptide. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for none of the compounds. In the present case, the claim so lack support, and the application so lacks disclosure, that a meaningful search over the claimed scope is impossible.

Present claims 20, 22 and 23 relate to an agent defined in terms of the method for identifying it. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is not to be found, however, for such agent. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the claimed scope is impossible.

Present claims 24 and 26 relate to a compound defined by reference to a desirable characteristic or property, namely inhibiting the activity of the CPS1 polypeptide.

The claims cover all compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for none of such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the claimed scope impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/042444 A2(51) International Patent Classification⁷: **C12N 15/00**

(21) International Application Number: PCT/US01/43381

(22) International Filing Date:
21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,649 22 November 2000 (22.11.2000) US
60/252,732 22 November 2000 (22.11.2000) US(71) Applicants (for all designated States except US): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 (CH). **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; 20 Fomwood Drive, Suite 105, Ithaca, NY 14850 (US).

(71) Applicants and

(72) Inventors: **YODER, Olen** [US/US]; 4939 Concannon Court, San Diego, CA 91230 (US). **TURGEON, Barbara, G.** [CA/US]; 4939 Concannon Court, San Diego, CA 92130 (US). **LU, Shen-wen** [US/US]; 604 Winston Court, Apt. 4, Ithaca, NY 14850 (US).(74) Agent: **VIKSINS, Ann, S.**; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).**Published:**

— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:

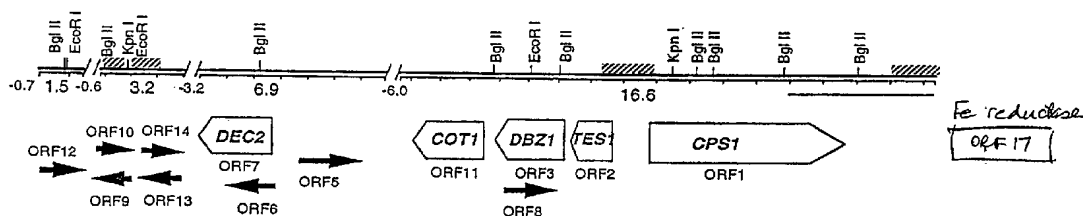
3 April 2003

(15) Information about Correction:

see PCT Gazette No. 14/2003 of 3 April 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: FUNGAL GENE CLUSTER ASSOCIATED WITH PATHOGENESIS



ORF 15 permease, MFS transporter
ORF 16 lactase precursor

(57) Abstract: Methods to identify orthologs of ungal CPS1 genes as well as fungal iron reductase and permease/and or MFS transporter genes, and uses thereof are provided.

WO 02/042444 A2